

TESIS DOCTORAL

Universidad Autónoma de Madrid
Facultad de Ciencias

**Programa de doctorado en Física de la
Materia Condensada, Nanociencia y Biofísica.**

INFLUENCIA DE LA DIVERSIDAD INICIAL Y LA PAUTA DE CAMBIO AMBIENTAL EN LA EVOLUCIÓN DEL BACTERIÓFAGO Q β

A7AAAAFFAAAFGG3EAFDDGB1F0B0CCHH
G7HH0>?/1FAFCGCGGHHFD1C<CCGG1<
FFFFFFBF:B:FEFBA9FBBFB9:BFFF/B
@M02255:285:000000000:C3NH3:1:21
GAATGTCGTGACATCTTCTCAACAGTC
GAACGAATGGCGGCCCTGTTAAGCTTTCG
CTACCTTTAGTTCGTTTAAACACGTTTGAT

Doctoranda
Pilar Somovilla Crespo

Directora de tesis
Ester Lázaro Lázaro

Madrid, 2021

TESIS DOCTORAL



EXCELENCIA
MARÍA
DE MAEZTU



Universidad Autónoma de Madrid
Facultad de Ciencias

Programa de Doctorado en Física de la
Materia Condensada, Nanociencia y Biofísica

INFLUENCIA DE LA DIVERSIDAD INICIAL Y LA PAUTA DE CAMBIO AMBIENTAL EN LA EVOLUCIÓN DEL BACTERIÓFAGO Q β

Doctoranda

Pilar Somovilla Crespo

Directora de tesis

Ester Lázaro Lázaro

Madrid 2021

Memoria presentada por la Licenciada en Ciencias Físicas D^a Pilar Somovilla Crespo, adscrita al Programa de Doctorado en Física de la Materia Condensada, Nanociencia y Biofísica, para optar al grado de Doctor por la Universidad Autónoma de Madrid.

El trabajo presentado en esta tesis doctoral ha sido llevado a cabo en el Centro Nacional de Biotecnología (CSIC) y el Centro de Astrobiología (CSIC-INTA), bajo la supervisión de la Dra. Ester Lázaro Lázaro y financiado con una ayuda, dentro del marco del Programa Estatal de Promoción del Talento y su Empleabilidad del Plan Estatal de Investigación Científica y Técnica y de Innovación 2013-2016, de un contrato predoctoral para la formación de doctores del Ministerio de Ciencia e Innovación (BES-2015-074460).

Diseño de portada e ilustraciones: Edson Ferrage

ERROR DE CÁLCULO

*“Y si después de todo...
resulta que las pistas
eran el tesoro.”*

Guille Galván

AGRADECIMIENTOS

¡Y aquí está mi tesis!

En el siguiente documento, Ester Lázaro y yo hemos intentado reflejar parte del trabajo que hemos realizado en estos últimos años. Han sido tiempos difíciles, y terminar de escribir una tesis en estas circunstancias, con una pandemia mundial en la que verse o abrazarse está prohibido y con este mundo tan loco... no era algo que fuera previsible y no ha sido fácil. En la vida hay caminos largos y carreras de fondo, y éste ha tenido de ambos elementos. Pero durante esta andadura he tenido a mi lado a personas a las que quiero agradecer todo lo que me han regalado, así que estas líneas van dedicadas a ellas.

En primer lugar quiero agradecer a mi familia. A mis padres por darme siempre la oportunidad de continuar con mis estudios y haberse esforzado tanto en ello. A mis hermanos (Julio, Bea, Sara, Jose y Víctor) por ser las personas que más me quieren en este mundo, las que más me cuidan y las que siempre voy a tener a mi lado. A Lucía, mi *Pequeña gran revolución*, por salvarme de este mundo de locos y cubrirme de alegría, inocencia pura y amor del de verdad. Aunque seamos de decirlo pocas veces, os quiero con locura.

En segundo lugar, querría hacer un especial agradecimiento a Ester Lázaro por la buena acogida que me brindó cuando llegué a su laboratorio. No eran las mejores circunstancias pero siempre me sentí bien recibida. Te doy las gracias por haber aprendido tanto de ti, por haber tenido tan buena disposición en enseñarme todo desde cero, repetirlo cien veces si era necesario (no es fácil enseñar a un físico biología molecular, y más experimental), y por haberme hecho disfrutar y apasionarme con la virología.

En tercer lugar, a todos los miembros del grupo, comenzando por Esther A., que fue quien me puso mi primera pipeta en la mano y me ayudó con mis primeros pasos dentro del laboratorio, a Daniel, a Ramón y a todos los estudiantes que han pasado por allí en estos años, y, cómo no, mis queridas Mara y Elena, con las que he compartido y aprendido mucho, no solo de biología. Gracias, sois fantásticas, gran calidad como científicas y una enorme calidad como personas. Tampoco podría dejar de mencionar al resto de Mis Chicas del Labo (Bea, Carol, Almudena y María). Gracias por hacer que todos los días fuera con una sonrisa. No he podido tener un mejor ambiente y unas mejores compañeras. Sois Luz.

También querría agradecer al Centro de Astrobiología por la oportunidad que me brindó de poder realizar mi tesis en el laboratorio de Evolución Molecular, así como a todas las personas con las que he compartido mi experiencia allí. Desde el primer momento me sentí muy acogida y arropada, y siempre he encontrado una mano amiga dispuesta a echarme un cable. Y hablando de manos amigas, quiero agradecer notablemente al grupo de Termo, ellos han sido esenciales en todo este proceso. Gracias a Elena, Bea y Carol por ser Oxo (Oxo is the best), y a Sara y a Héctor por todas las visitas al labo y por

todas las charlas post. Siempre seréis una gran parte de esto. Tampoco podría no mencionar de una manera especial a todos los que han pasado en algún momento por el Coffee Club (Jose, Santos, Alberto, Bruno, Maca, Laura, las Cris, Carmen, Juanma, Diego, Juanjo, Karol, los Javis, Esther... y un largo etcétera, que no se me ofenda nadie por no aparecer en la lista). Gracias por todas las charlas durante las comidas, todos los escapes room, todas las quedadas y todo el cariño.

Gracias a mi familia elegida, a mi Gentuza, pero adorable (Bea, Catia, Cynthia, Edson, Lu, Pipo, Rafa, Silvia y Víctor). Ellos son el cable a tierra y el hogar al que siempre regresar. A Guille, Carlos y Marian, por ser mi Consejo de Sabios y estar ahí a pesar de los años y los cruces de caminos. Gracias hermana por vivir como propias mis alegrías y mis penas.

A A Por Todo, por ayudarme tanto a crecer y a convertirme en la persona que soy ahora, por darme la oportunidad de ver otras realidades, otras formas de hacer las cosas y crear en mí una visión crítica, luchadora y participativa. A toda esa gente que, gracias a formar parte de esta gran familia, he conocido durante todos estos años, GRACIAS, a los que están cerca, a los que están lejos y a los que ya no están. *Para ti, cada mañana el Sol.*

Y finalmente, dar las gracias a los grupos de Susanna Manrubia y José Cuesta por brindarme la oportunidad de asistir a los seminarios y enseñarme tanto sobre el mundo de la ciencia.

Pilar Somovilla Crespo

ÍNDICE

1. RESUMEN	1
2. INTRODUCCIÓN	3
2.1 LA EXPERIMENTACIÓN EN EL LABORATORIO APLICADA AL ESTUDIO DE LA EVOLUCIÓN BIOLÓGICA.....	3
2.2 LOS VIRUS DE RNA COMO MODELO EVOLUTIVO	6
2.2.1 Tasa de error o tasa de mutación.....	6
2.2.2 Cuasiespecies virales	8
2.3 SEGUIMIENTO DE UN EXPERIMENTO DE EVOLUCIÓN CON VIRUS.....	10
2.4 EFECTO DE LA PAUTA DE CAMBIO DE LAS PRESIONES SELECTIVAS EN LA ADAPTACIÓN	12
2.5 EFECTO DE LA DIVERSIDAD PREEXISTENTE EN LA ADAPTACIÓN	14
2.6 EL BACTERIOFAGO Q β	16
2.6.1 Características generales.....	16
2.6.2 Infección viral.....	18
2.7 LA TEMPERATURA COMO PRESIÓN SELECTIVA EN LA ADAPTACION VIRAL	20
3. OBJETIVOS.....	23
4. MATERIALES Y MÉTODOS.....	25
4.1 MATERIALES	25
4.1.1 Medios de cultivo y crecimiento	25
4.1.2 Disoluciones y tampones.....	25
4.2 MÉTODOS EXPERIMENTALES, BIOINFORMÁTICOS Y TEÓRICOS	26
4.2.1 Virus y bacterias. Procedimientos generales de infección.....	26
4.2.2 Propagación del virus Q β _{WT}	27
4.2.3 Ensayos de determinación de <i>fitness</i>	29
4.2.4 Obtención de mutantes únicos mediante mutagénesis dirigida	30
4.2.5 Determinación de las secuencias consenso de las poblaciones de Q β	31
4.2.6 Procedimientos y análisis de secuenciación masiva	33
4.2.7 Estudio de la diversidad de las poblaciones virales.....	37

5. RESULTADOS	41
5.1 CAPÍTULO 1: Caracterización de poblaciones del bacteriófago Q β que difieren en el grado de diversidad genética.....	41
5.1.1 Análisis mediante NGS de las poblaciones Q β _{P2} y Q β _{P25}	42
5.1.2 Variación de los índices de diversidad en función del umbral de representación de los haplotipos.....	54
5.2 CAPÍTULO 2: Adaptación a temperatura elevada de las poblaciones del bacteriófago Q β que difieren en el grado de diversidad	59
5.2.1 Variación del <i>fitness</i> a lo largo de la adaptación	59
5.2.2 Dinámica de adquisición de mutaciones durante la adaptación a 43 °C.	61
5.2.3 Comparación de las mutaciones adquiridas en los dos conjuntos de líneas evolutivas	63
5.2.4 Convergencia evolutiva entre líneas	66
5.3 CAPÍTULO 3: Estudio mediante NGS de la influencia de la diversidad genética preexistente en la adaptación a 43 °C	69
5.3.1 Distribuciones de haplotipos en las líneas C43 _{P2} .3 y C43 _{P25} .1.....	69
5.3.2 Evolución de la frecuencia de las mutaciones más representadas a lo largo de la adaptación a 43 °C.....	72
5.3.3 Variaciones en los haplotipos más representados a lo largo de la adaptación.....	75
5.3.4 Determinación de la diversidad genética en las poblaciones C43 _{P2} .3 y C43 _{P25} .1 mediante la estimación de índices de diversidad.....	80
5.3.5 Variación de los índices de diversidad en función del umbral de representación de los haplotipos.....	84
5.4 CAPÍTULO 4: Influencia de la pauta de cambio en la adaptación a temperatura elevada.....	87
5.4.1 Variación del <i>fitness</i> a 43 °C a lo largo de la adaptación.....	88
5.4.2 Dinámica de adquisición de mutaciones en la secuencia consenso	90
5.4.3 Comparación de las mutaciones adquiridas en la secuencia consenso en función de la pauta de cambio de temperatura	92
5.4.4 Grado de convergencia durante la adaptación	95
5.5 CAPÍTULO 5: Estudio mediante NGS de la variación de la diversidad genética intrapoblacional en función de la pauta de cambio en las presiones selectivas	97
5.5.1 Distribuciones de haplotipos	97

5.5.2	Evolución de la frecuencia de las mutaciones más representadas a lo largo de la adaptación	100
5.5.3	Variaciones de los haplotipos más representados a lo largo de la adaptación.....	103
5.5.4	Determinación de la diversidad genética en las poblaciones G1 _{P25.2} y G2 _{P25.1} mediante la estimación de índices de diversidad	106
5.6	CAPÍTULO 6: Valor selectivo de las mutaciones seleccionadas durante la adaptación a temperatura elevada	111
5.6.1	Detección de mutaciones en la secuencia consenso en función de la temperatura de evolución.....	111
5.6.2	Efectos en el <i>fitness</i> de las mutaciones seleccionadas a temperatura elevada cuando están presentes en el contexto genético del virus salvaje	114
5.7	CAPÍTULO 7: Adaptación del bacteriófago Q β a temperaturas alternantes...117	
5.7.1	<i>Fitness</i> a 30 °C y 43 °C de las poblaciones evolucionadas de forma constante y alternante a esas temperaturas	117
5.7.2	Mutaciones adquiridas durante la evolución a temperaturas constantes y alternantes	119
5.7.3	Estudio comparativo de la diversidad genética contenida en las poblaciones evolucionadas a temperaturas constantes o alternantes	122
5.7.4	Variación de los índices de diversidad en función de la temperatura de evolución	128
6.	DISCUSIÓN	135
6.1	La propagación de un clon biológico de Q β en condiciones estándar aumenta la diversidad y complejidad de la población	135
6.2	Durante la propagación de Q β a 37 °C se genera un espectro de mutantes que contiene en bajas frecuencias mutaciones relevantes para la adaptación a 43 °C...137	
6.3	La diversidad genética generada durante la evolución a 37 °C acelera la adaptación a 43 °C.....	137
6.4	La adaptación a 43 °C a partir de una población ancestral más diversa produce un cuello de botella poblacional.....	139
6.5	El seguimiento de haplotipos a lo largo de la adaptación muestra la aparición de combinaciones de mutaciones que estaban presentes en genomas separados en la población ancestral	141

6.6	La velocidad de adaptación y las mutaciones seleccionadas a temperatura elevada dependen de la pauta de cambio.....	142
6.7	La línea G1 _{P25.2} presenta los valores mayores para los índices de diversidad basados en abundancia	145
6.8	La adaptación a condiciones de temperatura alternantes tiene lugar a través de las mutaciones que confieren ventajas frente a la presión selectiva más intensa	146
6.9	Análisis global de las mutaciones seleccionadas a temperatura elevada	147
6.10	Posibles efectos de las mutaciones seleccionadas a temperatura elevada....	150
7.	CONCLUSIONES	153
8.	REFERENCIAS	155
9.	ANEXOS	171
9.1	LISTA DE ABREVIATURAS.....	171
9.2	LISTA DE FIGURAS	173
9.3	LISTA DE TABLAS	176
9.4	PUBLICACIÓN.....	179

1. RESUMEN

La evolución viral está condicionada por multitud de factores entre los que se encuentran la velocidad de cambio en las presiones selectivas y la diversidad genética preexistente en las poblaciones. En general, cuanto más intenso sea el cambio ambiental mayor será la reducción del tamaño de la población, aumentando así la dificultad para generar y seleccionar mutaciones adaptativas, lo que puede conducir a la extinción. Por el contrario, los cambios graduales, que aumentan las presiones selectivas de forma progresiva, suelen ser más compatibles con la adaptación. La diversidad genética preexistente puede facilitar el proceso si en el espectro de mutantes que componen las poblaciones virales se encuentra ya algún genoma con ventajas adaptativas en las nuevas condiciones.

En esta tesis doctoral hemos planteado varios experimentos de evolución en los que hemos estudiado los cambios fenotípicos y genotípicos que experimenta un bacteriófago de RNA (Q β) cuando se propaga a temperaturas diferentes de la óptima, siguiendo diferentes patrones de cambio. A lo largo del proceso se han determinado los valores de la tasa de crecimiento, las secuencias consenso de las poblaciones y también se ha caracterizado el espectro de mutantes mediante tecnologías de secuenciación masiva.

Los resultados más relevantes obtenidos muestran que las poblaciones que contienen mayor diversidad genética inicial se adaptan más rápido a un aumento de temperatura (de 37 °C a 43 °C) que las que parten de una población menos diversa. Los análisis de secuenciación masiva indican que en el primer caso se produce un cuello de botella poblacional que podría ser debido a la rápida selección de algún genoma minoritario que estuviera ya presente en la población ancestral. La velocidad de la adaptación también aumenta con la velocidad de cambio de temperatura. A pesar de que hay bastante coincidencia entre las mutaciones que se seleccionan al final del proceso en todos los casos analizados, las vías adaptativas seguidas no son las mismas, como se demuestra por las diferencias en las mutaciones que se fijan en primer lugar y el número de mutaciones de la población ancestral que se mantienen en el nuevo ambiente. La estructura de la población y la diversidad genética contenida en ella también experimentan cambios de distinta intensidad en función de la pauta de cambio y el estado inicial de la población.

Por último, la propagación de Q β en condiciones que alternan la replicación a dos temperaturas subóptimas favorece la adaptación a la que genera la presión selectiva más intensa y lo hace a través de las mismas mutaciones que se seleccionan durante la adaptación constante a esa misma temperatura. Sin embargo, la estructura de la población alternante se asemeja más a la evolucionada de forma constante a la temperatura que ejerce menor presión selectiva, mostrando así que ambas temperaturas influyen en las propiedades de la población evolucionada.

2. INTRODUCCIÓN

2.1 LA EXPERIMENTACIÓN EN EL LABORATORIO APLICADA AL ESTUDIO DE LA EVOLUCIÓN BIOLÓGICA

Desde que tenemos constancia de ello, gran parte de lo que nos rodea está sufriendo cambios de manera constante, y, desde sus orígenes, el hombre ha intentado entender la dinámica y las razones subyacentes a esos cambios. La evolución puede considerarse como el conjunto de procesos de transformación de la materia desde el origen del Universo hasta nuestros días, englobando tanto la evolución de la materia inorgánica como la de la materia orgánica. De esta última, se deben tener en cuenta dos grandes procesos: la etapa del origen de la vida y la etapa de la evolución biológica. El origen de la vida comprende el conjunto de procesos químicos y físicos que posibilitaron la transformación de la materia inorgánica en orgánica y de esta en los primeros seres vivos. La evolución biológica propiamente dicha comprende el conjunto de procesos relativos a la herencia de los caracteres biológicos, desde los organismos vivos más sencillos a los más complejos.

Tradicionalmente, el estudio de la evolución se ha abordado realizando análisis comparativos entre especies, teniendo en cuenta características morfológicas y fisiológicas y, más recientemente, características moleculares, sobre todo secuencias de genes o de proteínas. Estos estudios han resultado muy útiles para establecer relaciones filogenéticas entre especies. En las últimas décadas, con los avances en los métodos de secuenciación, que permiten determinar incluso la secuencia de genomas completos, el poder resolutivo de este tipo de comparaciones ha aumentado considerablemente, permitiendo obtener cronologías muy precisas sobre cómo las especies que han poblado nuestro planeta se han ido sucediendo (Brockhurst y col., 2011) (Fig. 1). Sin embargo, estos estudios no nos permiten discernir cómo actúa la evolución ni cuáles son los principios que rigen la transformación de unas especies en otras. Así, existen muchas preguntas, aparentemente sencillas, que todavía requieren una respuesta. ¿Cómo interaccionan las mutaciones entre sí? ¿Cuál es la influencia del ambiente en el valor selectivo de las mutaciones? ¿Cuál es la contribución relativa de la selección natural y la deriva génica? ¿Cómo influye la pauta de cambio de las presiones selectivas en las vías adaptativas seguidas por las poblaciones? ¿Qué tipo de restricciones existen en la optimización de determinados rasgos? Todas estas cuestiones no pueden ser respondidas utilizando únicamente comparaciones de genes, sino que requieren un estudio del proceso evolutivo en directo, algo difícil de conseguir cuando se observa la naturaleza.

Por un lado, está el problema de las largas escalas de tiempo que normalmente se requieren para que los resultados de la evolución se manifiesten. Por otro, la dificultad añadida de que las poblaciones naturales evolucionan en entornos complejos donde múltiples variables, que no podemos controlar, actúan de forma interdependiente y experimentando a menudo patrones de cambio no uniformes. En estas condiciones, establecer relaciones entre las variables ambientales, los cambios en las propiedades de las poblaciones y la presencia de mutaciones específicas es un gran reto, existiendo,

además, muchas incógnitas sobre la influencia de la historia evolutiva y el azar en el estado evolutivo final.

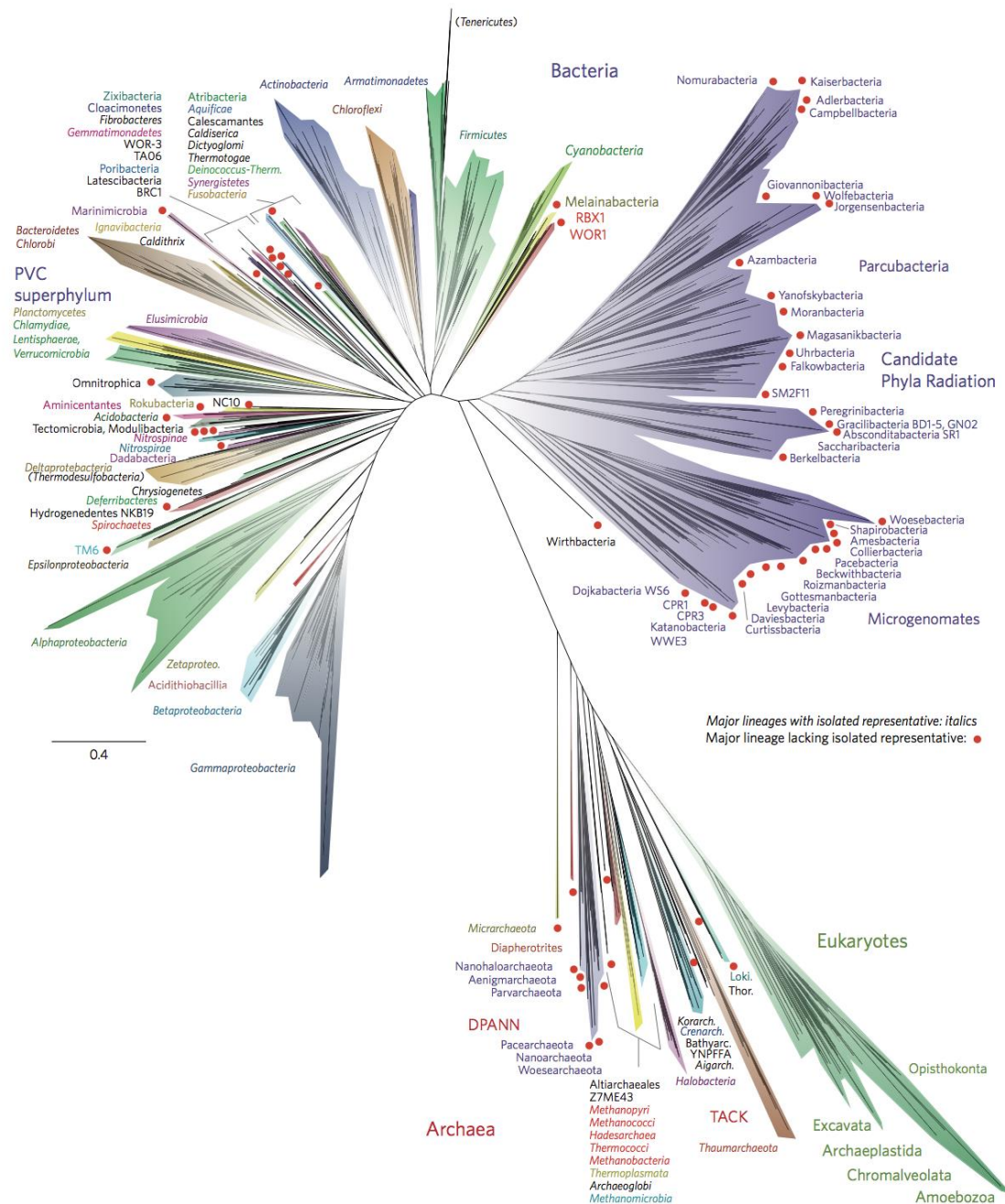


Figura 1: Una vista actual del árbol de la vida, que abarca la diversidad total representada por los genomas secuenciados. (Imagen: Hug y col., 2016).

Todos estos problemas han intentado ser solventados, al menos en parte, mediante la realización de experimentos en el laboratorio, en condiciones controladas por el experimentador, que permitan establecer correlaciones entre cambios en las variables

y adaptaciones específicas. Las poblaciones objeto de estudio tienen que evolucionar de forma rápida y, a ser posible, deben ser fáciles de manejar. En este sentido, los microorganismos (virus y bacterias), con sus cortos tiempos de generación, sus altas tasas de error y el elevado tamaño de sus poblaciones, son capaces de, en un corto espacio de tiempo, generar poblaciones muy diversas que facilitan y aceleran la adaptación, permitiendo en muchos casos observar este proceso en tiempo real (Elena y Lenski, 2003; Buckling y col., 2009).

Aunque la mayoría de los experimentos de evolución no pueden tener en cuenta la complejidad del mundo real, poseen muchas ventajas que los convierten en un interesante complemento a los estudios realizados en condiciones naturales. Por ejemplo, el hecho de poder realizar múltiples réplicas de un mismo experimento permite observar el papel jugado por el azar en el resultado final, así como la reproducibilidad de la evolución (Gould, 1989). La dinámica evolutiva puede ser observada en tiempo real y, además, es posible conservar un registro fósil de todos los estados por los que ha pasado una población a lo largo de la historia evolutiva recreada en el laboratorio. Finalmente, además de los parámetros ambientales, hay otros, como el tamaño poblacional (de Visser y col., 2005; Desai y col., 2007) y la tasa de error, que también puede ser manipulados fácilmente, algo de mucha ayuda a la hora de intentar discernir principios evolutivos básicos.

En general, los estudios de evolución experimental parten de una población ancestral bien caracterizada que se propaga durante un número determinado de generaciones bajo ciertas condiciones. Durante el proceso se producen mutaciones, algunas de ellas ventajosas, de modo que los individuos portadores de estas últimas pasarán a estar más representados en la población (Lang y Desai, 2014). Con el tiempo, la acción conjunta de la mutación y la selección natural conducirán a la generación de una población optimizada en las condiciones en que se ha realizado el experimento. Esta población evolucionada podrá ser analizada y comparada con la ancestral para determinar los cambios fenotípicos (Gallet y col., 2012) y genéticos experimentados. En el caso de estos últimos se puede hacer un seguimiento mediante tecnologías clásicas tipo Sanger (Bull y col., 1997; Wichman y col., 1999; Bollback y Huelsenbeck, 2007; 2009; Cabanillas y col., 2013), o emplear métodos de secuenciación masiva que permiten identificar la mayor parte de las mutaciones beneficiosas que están contribuyendo a la adaptación (Brockhurst y col., 2011; Dettman y col., 2012; Lang y Desai, 2014; Bruger y Mars, 2018).

El estudio de evolución experimental más largo, y probablemente el más famoso, consistió en la propagación de doce poblaciones de *Escherichia coli* que fueron pasadas diariamente durante más de 68000 generaciones (Lenski y col., 1991; Lenski, 2017) y, a día de hoy, aún continúan con su proceso de optimización evolutiva (Good y col., 2017). Otros estudios de evolución llevados a cabo con bacterias han estudiado la adaptación a temperaturas elevadas (Tenaillon y col., 2012), el cambio en la fuente de carbono utilizada como nutriente (Blount y col., 2012), la emergencia de la resistencia a antibióticos (Yen y Papin, 2017) o el uso de la heterogeneidad clonal en su beneficio (Beaumont y col., 2009).

En esta tesis doctoral se prestará especial atención a los estudios de evolución experimental llevados a cabo con virus, en particular, a los que tienen un genoma de

RNA que, como se verá a continuación, constituyen el sistema biológico que evoluciona más rápido en la naturaleza. Además de contribuir al conocimiento del proceso evolutivo, este tipo de estudios tienen también un componente aplicado, debido a que gran parte de la dificultad para el tratamiento de las enfermedades causadas por virus provienen del alto potencial evolutivo de estas entidades biológicas (Domingo y col., 2012; 2019; Perales y Domingo, 2016). La infección de nuevos huéspedes, la adaptación a drogas antivirales o la aparición de variantes que no pueden ser neutralizadas por los anticuerpos son solo algunos ejemplos de cómo la gran adaptabilidad de los virus de RNA dificulta el hallazgo de terapias efectivas.

Mientras que el uso de virus de RNA patógenos puede ser una dificultad para la realización de estudios básicos de evolución, los bacteriófagos (virus que infectan bacterias) pueden ser un modelo más apropiado, debido a la carencia de medidas especiales de seguridad para su manejo. Aunque la mayoría de los bacteriófagos caracterizados hasta la fecha poseen un genoma de DNA, los pertenecientes a la familia Leviviridae (Olsthoorn y col., 2011; Callanan y col., 2018) tienen un genoma de RNA de polaridad positiva, lo que los convierte en un buen sistema para estudiar los procesos evolutivos que ocurren en otros virus con genomas similares.

2.2 LOS VIRUS DE RNA COMO MODELO EVOLUTIVO

Los virus y los viroides son los sistemas replicativos más simples que existen en la naturaleza. Ambos necesitan una célula hospedadora en la que introducirse para tener así disponibles las estructuras y las moléculas necesarias para completar su ciclo vital. Son, por tanto, parásitos intracelulares. Casi todas las especies que existen en la biosfera son susceptibles de ser infectadas por algún tipo de virus (Murphy, 1996), algo que está relacionado con la gran diversidad de estrategias que estos poseen para almacenar su información genética (DNA o RNA, de cadena doble o sencilla, circular o lineal) y llevar a cabo su replicación.

2.2.1 Tasa de error o tasa de mutación

Quizás la característica más importante de los virus de RNA es la elevada diversidad genética contenida en sus poblaciones. Este hecho se debe en gran medida a la alta tasa de error que presentan durante su replicación. La tasa de error o tasa de mutación de un virus se define como la frecuencia de incorporación de nucleótidos erróneos durante la copia de su genoma. Esta tendencia a equivocarse determina la capacidad del virus para hacer frente a los cambios ambientales, al mismo tiempo que mantiene la información esencial (Domingo y Holland, 1994; 1997; Holland y col., 1992; Holland, 1993). Las tasas de error estimadas para los virus RNA se encuentran entre 10^{-3} a 10^{-5} sustituciones por nucleótido y ronda de replicación (s/n/r) (Batschelet y col., 1976; Drake y Holland, 1999; Domingo, 2007; Sanjuán y col., 2010), varios ordenes de magnitud mayores que las de los virus de DNA (10^{-8} – 10^{-10} s/n/r) (Drake, 1991; Cuevas

y col., 2009) o que las de DNA celular ($10^{-10} - 10^{-11}$ s/n/r) (Kunkel y Alexander, 1986; Echols y Goodman, 1991).

Puesto que el tamaño genómico típico de un virus de RNA está en torno a los 10^4 nucleótidos, las tasas de error que se han indicado suponen que la mayor parte de los nuevos genomas generados en cada replicación difiere en una o varias mutaciones del genoma parental. De esta alta tasa de mutación de los virus RNA en cierta medida es responsable la falta de actividades correctoras de las enzimas que copian el RNA viral (las RNA replicasas y las transcriptasas reversas), que sí que están presentes en las DNA polimerasas (Bernard y col., 1989). La ausencia de mecanismos de reparación posteriores a la replicación (Friedberg y col., 2006), las elevadas tasas de recombinación genética de algunos virus (Mikkelsen y Pedersen, 2000; Domingo, 2007; Domingo y col., 2008) y los reordenamiento genómicos en aquellos que poseen genomas segmentados (Webster, 1999) son factores adicionales que contribuyen aún más a aumentar la diversidad genética en las poblaciones de virus RNA.

Una limitación evolutiva que ocurre en todos los tipos de organismos es que el máximo tamaño genómico que puede ser mantenido sin la pérdida de la información genética está inversamente correlacionado con la tasa de error de la replicación (Drake, 1991). Esta restricción impone un límite al tamaño permitido para los genomas de los virus de RNA, que en el caso de los bacteriófagos está en el rango de 3 a 15 Kb (Gago y col., 2009). Como consecuencia, la información genética debe estar altamente compactada, siendo frecuentes las proteínas multifuncionales y las fases de lectura solapantes que también pueden incluir regiones regulatorias. Además, la estructura del RNA también es funcional, puesto que participa en su propia replicación y en la interacción con componentes virales y celulares. Es fácil entender que este tipo de genomas son muy susceptibles al efecto de las mutaciones, las cuales suelen tener efectos deletéreos (Domingo-Calap y col., 2009; Sanjuán, 2010; Vale y col., 2012) y son más propensas a interactuar entre ellas que en genomas menos compactos (Sanjuán y col., 2004; Elena y col., 2010).

La principal consecuencia de las altas tasas de error de los virus de RNA es que dentro de sus poblaciones coexisten una amplia variedad de mutantes. Estos mutantes difieren en sus secuencias genómicas, en la naturaleza de los productos que codifican y en sus valores de eficacia biológica o *fitness*. El *fitness* de un virus puede definirse como su capacidad para originar descendientes viables en un ambiente concreto (Domingo y Holland, 1997). Muchas veces el *fitness* se equipara a la tasa de crecimiento, pero en realidad es un carácter mucho más complejo, ya que depende de más factores (Manrubia y Lázaro, 2006). Entre ellos están la estabilidad ambiental y las interacciones que se establecen con el resto de mutantes de la población. Por esta razón, los valores de *fitness* son siempre relativos y se refieren a unas condiciones concretas (Quiñones-Mateu y Arts, 2006).

La acción de la selección natural sobre la amplia diversidad de mutantes que contienen las poblaciones virales determina que estos se distribuyan con diferentes frecuencias. La selección positiva es el proceso por el que un mutante ventajoso aumenta su frecuencia debido a su mayor capacidad para originar descendientes que el resto. Por el

contrario, la selección negativa disminuiría las frecuencias de los mutantes de menor *fitness*, pudiendo incluso llegar a eliminarlos. Sin embargo, las altas tasas de error de la replicación de los virus RNA, determinan que aquel que tiene mayor *fitness* nunca va a desplazar totalmente a los restantes. La continua producción de mutaciones va a favorecer la presencia, aunque sea en baja concentración, de un conjunto muy extenso de variantes, que se conoce como espectro de mutantes, y que va a ser responsable de gran parte de las propiedades evolutivas de la población.

2.2.2 Cuasiespecies virales

Las complejas distribuciones de mutantes que forman las poblaciones virales a menudo se conocen con el nombre de cuasiespecies (Biebricher y Eigen, 2006). El término cuasiespecies fue inicialmente propuesto por Eigen y Schuster para describir la dinámica evolutiva de las poblaciones de replicones primitivos que poblaron la Tierra antes del origen de la vida celular, según dicta la teoría del mundo RNA (Eigen, 1971; Eigen y Schuster, 1977). De acuerdo con la teoría propuesta por Eigen, dado el suficiente tiempo de evolución en un ambiente constante, la acción combinada de la selección natural y las elevadas tasas de error actuando sobre una población infinita de replicadores (cada uno de ellos definido por un valor de *fitness*) daría lugar a un estado estacionario en el que cada mutante ocuparía una frecuencia constante dentro del conjunto de la población.

Aplicado a la virología, el término cuasiespecies hace referencia a un conjunto de variantes genéticas no idénticas, relacionadas a través de mutaciones, y que están sujetas a un proceso continuo de variación, competición y selección (Domingo y col., 2012; Andino y Domingo, 2015; Domingo y Perales, 2019). Hay varios hechos que diferencian las cuasiespecies virales de las cuasiespecies moleculares definidas de forma teórica. Las poblaciones virales evolucionan en ambientes complejos, en los que las variables están continuamente sujetas a cambio, lo que dificulta llegar a alcanzar el equilibrio mutación-selección (estado estacionario) descrito para las cuasiespecies moleculares. Las poblaciones virales tampoco son infinitas y, además, experimentan frecuentes cuellos de botella poblacionales que reducen la diversidad genética e introducen aleatoriedad en el sistema. Finalmente, otro elemento que no fue considerado en la teoría de cuasiespecies en su formulación inicial es la existencia de una amplia variedad de interacciones, tanto entre mutantes como entre mutaciones. En una población de virus RNA, las distintas variantes interactúan a nivel funcional y contribuyen de manera colectiva a sus propiedades (Lauring y Andino, 2010), ayudando a mantener la eficacia biológica a pesar de la continua producción de mutaciones (Eigen, 1993).

El bacteriófago Q β , el virus de la fiebre aftosa (VFA) y el virus de la estomatitis vesicular (VSV) (Domingo y col., 1978; 1980; Holland y col., 1982; Sobrino y col., 1983) representaron los primeros hallazgos experimentales de cuasiespecies en virus. Más tarde este concepto se generalizó a otros virus de RNA y DNA y otros microorganismos (Webb y Blaser, 2002; Solé y Deisboeck, 2004; Ge y col., 2007; Fox y Loeb, 2010;

Weissmann y col., 2011; Napoletani y col., 2013; Amor y Solé, 2014; Chevereau y col., 2015; Wagner y col., 2016).

Las poblaciones de virus RNA pueden caracterizarse por su secuencia consenso, que es la determinada por los nucleótidos más representados en cada una de las posiciones del genoma (Fig. 2). El espectro de mutantes está compuesto por todos los genomas que componen la cuasiespecie y constituye un reservorio de variantes que podrán ser utilizadas durante las adaptaciones futuras. La secuencia consenso es muy útil para identificar cambios mayoritarios que pueden ocurrir en la población, como los determinados por los cuellos de botella poblacionales o los cambios ambientales que favorecen a unos mutantes sobre otros. Sin embargo, conocer la secuencia consenso no informa sobre la composición interna de la población, es decir su espectro de mutantes, que es donde reside su potencial evolutivo. La misma secuencia consenso puede obtenerse para diferentes distribuciones de mutantes, lo que muestra las aplicaciones limitadas de este concepto.

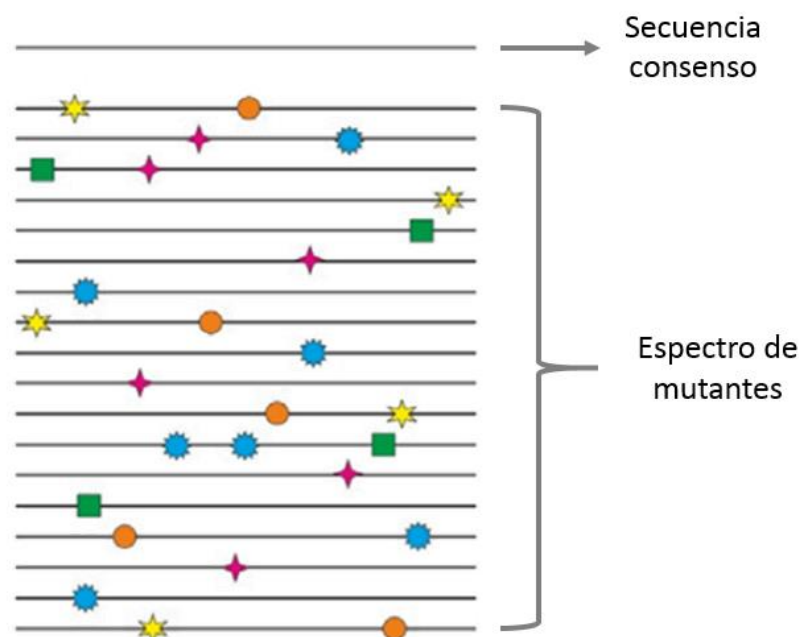


Figura 2: Esquema de una cuasiespecie. La secuencia consenso de la población (primera línea) vendrá definida por el valor promedio del conjunto de mutantes (resto de líneas). Los mutantes pueden contener mutaciones y la secuencia consenso puede no estar incluida en el conjunto de las secuencias de la cuasiespecie.

El hecho de que las poblaciones virales tengan tamaños finitos y que la mayoría de las mutaciones tengan efectos negativos limita la extensión de diversidad genética que pueden contener. La amplitud de un espectro de mutantes puede ser visualizada como la fracción ocupada dentro del espacio de secuencias, definido como todas las posibles variantes que se pueden generar a partir de una secuencia genómica. Cuando a cada genoma se le asigna un valor de *fitness* surgen los paisajes de *fitness* (Wright, 1931) (Fig. 3), una representación metafórica de las poblaciones, que permite visualizar los

movimientos de los componentes de una población a lo largo de la adaptación (Lauring y Andino, 2010; Acevedo y col., 2014; Schuster, 2016).

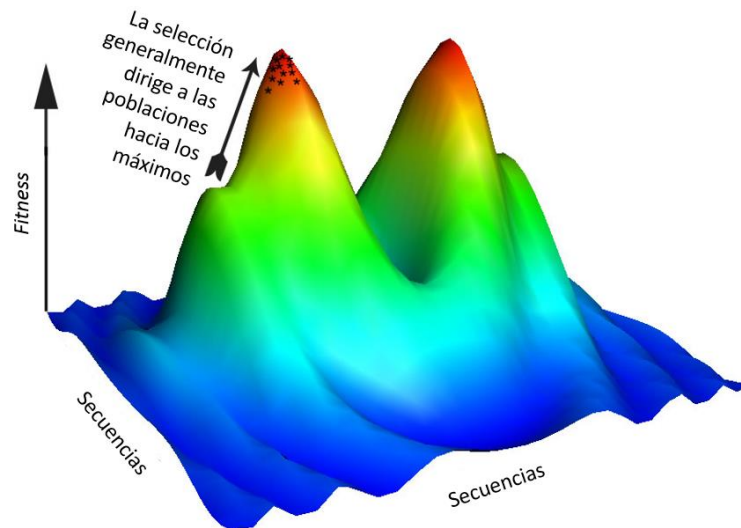


Figura 3: Paisaje de *fitness*. Representación esquematizada de la distribución de valores de *fitness* de una población viral. Adaptación de la imagen de © Laurence Loewe, 2016.

2.3 SEGUIMIENTO DE UN EXPERIMENTO DE EVOLUCIÓN CON VIRUS

Como ya se ha indicado, cualquier experimento de evolución supone la propagación de la población a estudiar durante cierto número de generaciones (Fig. 4). Un protocolo muy común para ello es el sistema de pases seriados, que, en el caso de las poblaciones virales, consiste en iniciar la infección del hospedador susceptible y permitir la replicación del virus durante un tiempo determinado. Esto es lo que se considera un pase. Una fracción de los virus producidos en cada pase es utilizada para iniciar el siguiente, y así sucesivamente.

Un diseño tan simple como el expuesto arriba permite manipular un número muy elevado de variables. Algunas de ellas afectan a la replicación del virus en el interior celular, como la temperatura, el medio de cultivo utilizado, la presencia de inhibidores, etc. Otras pueden afectar a su estabilidad en el medio extracelular, como la radiación ultravioleta, los pHs no óptimos, las condiciones de salinidad o, nuevamente, los cambios de temperatura. La estructura del medio en el que tiene lugar la replicación va a influir en la capacidad de difusión y la competición entre variantes. El tamaño poblacional, que en este caso correspondería al número de unidades infecciosas que se transfieren en cada pase, condiciona la extensión de diversidad genética que se expone a la acción de la selección natural. Otras variables que se pueden modificar fácilmente incluyen la densidad de hospedadores, el hospedador utilizado, la tasa de error (mediante el uso de agentes mutagénicos) o la condición inicial, es decir, las mutaciones que contiene la población al inicio de su propagación.

A lo largo de un experimento de evolución, es común determinar las variaciones de *fitness* y las mutaciones que han sido adquiridas, tanto en la secuencia consenso como en los genomas que componen el espectro de mutantes. Como se ha indicado, los valores de *fitness* pueden determinarse mediante la determinación de la tasa de crecimiento del virus (Bull y col., 2004) o mediante ensayos de competición con un virus de referencia (Quiñones-Mateu y Arts, 2006). En el último caso, la progenie viral se transfiere varias veces y, en cada pase, se determina la proporción del virus problema respecto al utilizado como referencia. La representación de esa proporción en función del tiempo suele poder ajustarse a una recta cuya pendiente sería el *fitness* del virus que se está analizando.

Para determinar las mutaciones que podrían ser responsables de los cambios de *fitness* que han podido ocurrir a lo largo del experimento de evolución, uno de los métodos es analizar la secuencia consenso mediante secuenciación Sanger. El problema es que solo las mutaciones presentes en cierta proporción de los genomas pueden ser detectadas de este modo. Conocer la diversidad de la población requiere necesariamente secuenciar una fracción representativa del conjunto de genomas que componen el espectro de mutantes. Una forma sencilla de conseguirlo consiste en aislar un número de placas líticas y determinar la secuencia consenso de la población viral contenida en cada placa (Bull y col., 1997; Wichman y col., 1999; Bollback y Huelsenbeck, 2007; 2009; Cabanillas y col., 2013). Dado que, en las condiciones usadas normalmente en el plaqueo, cada placa de lisis procede de la replicación de un único virus durante unas pocas rondas de replicación, puede considerarse un clon biológico cuya secuencia consenso sería igual a la del virus que inició el desarrollo de la placa.

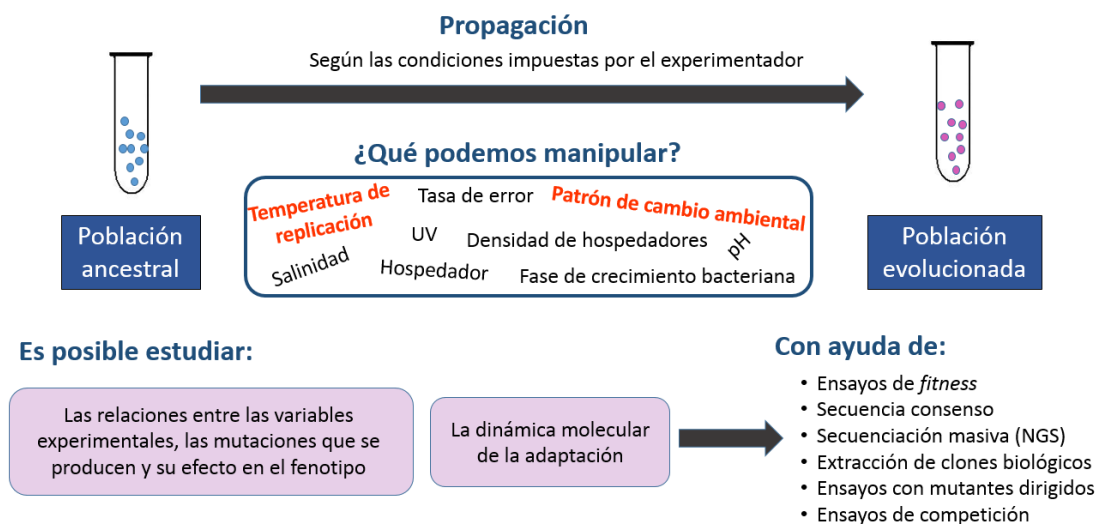


Figura 4: Esquema general de un ensayo de evolución experimental.

Las técnicas de secuenciación de nueva generación (Next generation sequencing, NGS), o también llamadas técnicas de secuenciación masiva, permiten secuenciar cientos de miles de individuos de una misma población, obteniéndose así una gran muestra de

genomas de su espectro de mutantes. Con esta amplia representación de secuencias, se pueden estudiar las frecuencias de cada polimorfismo y su dinámica a lo largo del tiempo (Brockhurst y col., 2011; Dettman y col., 2012; Lang y Desai, 2014; Bruger y Mars, 2018). El mayor problema de esta metodología es que solo permite determinar con fiabilidad la secuencia de fragmentos cortos, lo que dificulta la reconstrucción de haplotipos y, con ello, la identificación de las asociaciones entre mutaciones. Dado que en los virus de RNA es frecuente que el efecto de las mutaciones dependa del contexto genómico, algo que se conoce como epistasis (Fisher, 1918; Wolf y col., 2000; Burch y Chao, 2004; Phillips, 2008; Lázaro, 2014), se hace necesario desarrollar métodos fiables que permitan determinar cómo las mutaciones se combinan en los distintos genomas a lo largo de la adaptación.

2.4 EFECTO DE LA PAUTA DE CAMBIO DE LAS PRESIONES SELECTIVAS EN LA ADAPTACIÓN

Muchos estudios de evolución se centran en la dinámica adaptativa seguida por una población frente al cambio en una condición ambiental que se mantiene constante a lo largo del tiempo (Elena y Sanjuán, 2007; Collins, 2011; Barrick y Lenski, 2013). Sin embargo, los ambientes naturales rara vez cambian de esa manera tan simple. A menudo, los cambios ocurren gradualmente y, en ocasiones, experimentan fluctuaciones que se superponen a la tendencia principal. Otras veces las fluctuaciones no siguen ninguna tendencia o alternan dos valores o rango de valores.

La tasa de cambio de las presiones selectivas puede tener profundas consecuencias en la probabilidad de adaptación o extinción de las poblaciones, en las vías adaptativas seguidas y en el grado máximo de adaptación alcanzado en el proceso (Collins y de Meaux, 2009; Toprak y col., 2011; Gonzalez y Bell, 2013; Lindsey y col., 2013; Morley y col., 2015; Gorter y col., 2016; 2017; Morley y Turner, 2017).

La evolución en entornos que cambian gradualmente se ha modelado de forma teórica como un proceso de adaptación a un estado óptimo que se desplaza con el tiempo (Bello y Waxman, 2006; Kopp y Hermisson, 2007; 2009 a, b). Cuanto menor es la tasa de cambio, menor es la caída del *fitness* que la población experimenta en cada uno de los ambientes a los que se expone. En estas condiciones, las presiones selectivas tienen baja intensidad y, previsiblemente, la adaptación ocurrirá a través de la selección de mutaciones beneficiosas que proporcionan pequeños efectos sobre el *fitness* (Bello y Waxman, 2006; Kopp y Hermisson, 2007; 2009 a, b). Puesto que estas mutaciones son más abundantes que las de efecto mayor (Rokyta y col., 2005; Eyre-Walker y Keightley, 2007; Bons y col., 2018), se supone que habrá mayor diversidad de vías accesibles para la adaptación a una presión selectiva que cambia de forma gradual que para la adaptación a un cambio brusco en la misma condición ambiental. Además, debido a que las interacciones epistáticas entre mutaciones de pequeño efecto son más débiles (Schenk y col., 2013; Schoustra y col., 2016), también se ha sugerido que los patrones graduales permiten a las poblaciones alcanzar valores de *fitness* más altos.

Por el contrario, los cambios ambientales rápidos generalmente producen fuertes caídas de *fitness* que disminuyen severamente el tamaño de la población y, por tanto, la diversidad genética contenida en ellas (Bürger y Lynch, 1995; Bell y Collins, 2008; Gomulkiewicz y Houle, 2009; Kopp y Matuszewski, 2014). Este hecho limita la disponibilidad de mutaciones que podrían hacer posible la adaptación, a menos que esos mutantes ya existieran en la población. En ese caso el cuello de botella poblacional que se produciría tras el cambio de ambiente, tendría consigo la eliminación de los genomas que no facilitan (o incluso dificultan) la adaptación y el enriquecimiento de la población en los genomas conteniendo mutaciones con efecto positivo. El éxito de la adaptación dependerá de la frecuencia de esos genomas en la población inicial y del tamaño poblacional empleado durante la transmisión del virus, ya que si este es pequeño, los mutantes beneficiosos podrían ser eliminados por deriva génica.

En el caso de los cambios ambientales bruscos, la situación es aún más severa cuando se requiere una combinación particular de mutaciones para que la adaptación sea posible. Debido a la cantidad generalmente menor de soluciones adaptativas a los cambios bruscos, la evolución estaría más restringida bajo esta condición que cuando los cambios tienen lugar a un ritmo más lento (Yubero y col., 2017).

Hay muchos estudios experimentales, llevados a cabo con una variedad de sistemas biológicos, como bacterias (Perron y col., 2008; Lindsey y col., 2013), algas unicelulares (Collins y de Meaux, 2009), levaduras (Bell y Gonzalez, 2011; Gorter y col., 2016; Gorter y col., 2017) o virus (Morley y col., 2015; Hao y col., 2015; Morley y Turner 2017) que han explorado la influencia de la tasa de cambio ambiental en la adaptación. Las presiones selectivas ensayadas también son muy diversas e incluyen antibióticos (Perron y col., 2008; Lindsey y col., 2013), metales pesados (Gorter y col., 2016; 2017), nuevos anfitriones (Morley y col., 2015; Morley y Turner, 2017), estrés salino (Bell y Gonzalez, 2011) o cambios de temperatura (Hao y col., 2015).

De acuerdo con las predicciones de los modelos, en algunos estudios la tasa de extinción de la población fue mayor bajo cambios ambientales más rápidos (Perron y col., 2008; Bell y Gonzalez, 2011; Lindsey y col., 2013; Hao y col., 2015), debido a la fuerte disminución del tamaño de la población, y también, en algunos casos, porque los genotipos adaptados solo eran accesibles en una particular secuencia de ambientes. Este tipo de contingencias históricas pueden ocurrir cuando ciertas mutaciones que permiten la supervivencia a alta intensidad de la presión selectiva son perjudiciales en ausencia de otras mutaciones que solo pueden seleccionarse a una intensidad menor (Lindsey y col., 2013; Kashiwagi y col., 2014). Existen también ejemplos que muestran que las poblaciones que evolucionan bajo cambios graduales alcanzan valores mayores de *fitness* (Collins y de Meaux, 2009; Bell y Collins, 2008; Morley y col., 2015) y/o muestran más diversidad genética entre linajes replicados que aquellos que evolucionan bajo cambios más abruptos (Collins y de Meaux, 2009; Lindsey y col., 2013).

Por el contrario, los resultados obtenidos en otros estudios, mostraron ciertas discrepancias con las predicciones de los modelos, mostrando que los fenotipos obtenidos durante la adaptación a un cambio brusco tenían *fitness* más alto o contenían mayor diversidad genética que los obtenidos durante la adaptación a cambios más graduales. Finalmente, otros estudios no encontraron diferencias significativas ni en los

valores de *fitness* ni en el grado de diversidad contenido en las poblaciones que pudieran ser atribuidas a la velocidad de cambio ambiental (Gorter y col., 2016; 2017; Singhal y col., 2017).

La conclusión es que las diferencias observadas en las soluciones adaptativas en función de la pauta de cambio ambiental dependen de la presión selectiva y el sistema biológico utilizados. Los valores de las tasas de mutación y replicación, la distribución de los efectos de las mutaciones y la relevancia de las interacciones epistáticas en una particular condición ambiental, todo ello son factores que interaccionan entre sí y que van a influir en las posibilidades adaptativas de las poblaciones.

2.5 EFECTO DE LA DIVERSIDAD PREEXISTENTE EN LA ADAPTACIÓN

Una pregunta clave en evolución se refiere a cómo influye la diversidad genética contenida en las poblaciones en la adaptación a una nueva presión selectiva (Barrett y Schluter, 2008). Si la población contiene algún mutante que ya es beneficioso en las nuevas condiciones, ese mutante podría ser seleccionado, sin necesidad de tener que esperar a la generación de nuevas mutaciones. El hecho de que el efecto de las mutaciones dependa del ambiente hace que esta cuestión no sea trivial y que la historia evolutiva previa tenga una gran relevancia. Los estudios de genética cuantitativa llevados a cabo a lo largo de más de un siglo predicen que, a corto plazo, la selección natural actúa preferentemente sobre la variación genética preexistente, mientras que, a largo plazo, las mutaciones generadas *de novo* es cuando adquieren una importancia mayor (Hill, 1982; Roff, 1997).

La gran heterogeneidad de las poblaciones de los virus de RNA tiene como consecuencia que estas contengan un gran número de mutantes minoritarios con efectos neutrales o incluso deletéreos, pero que podrían ser beneficiosos cuando la población se enfrenta a una nueva presión selectiva, acelerando de este modo la adaptación y condicionando las vías adaptativas seguidas. Hay muchos ejemplos en los que se ha demostrado que la selección natural actúa sobre mutantes preexistentes en la población (Quakkelaar y col., 2007; Pennings, 2012; Borucki y col., 2013; Kutnjak y col., 2017). De hecho, algunos estudios muestran la existencia de un tipo particular de memoria en las poblaciones virales (Briones y col., 2003; 2006). Esta memoria consiste en el mantenimiento en bajas frecuencias de los mutantes con ventajas selectivas en los distintos ambientes por los que ha pasado la población. Esta especie de registro de lo que fue bueno en el pasado acelera la adaptación cuando la población se enfrenta a una circunstancia similar.

La probabilidad de fijación de las mutaciones beneficiosas preexistentes en una población depende de la magnitud del efecto beneficioso que confieren en el nuevo ambiente y del tamaño poblacional, es decir, del número de individuos que va a dar lugar a la siguiente generación. Cuanto mayor sean ambos parámetros menor va a ser la probabilidad de que las mutaciones se pierdan por deriva génica. En cualquier caso, para un tamaño poblacional dado, la pérdida de una mutación beneficiosa va a ser más alta cuando esta se genera *de novo* que cuando ya está presente en la población, ya que en el primer caso la mutación está presente inicialmente en un único genoma, mientras

que en el segundo es más probable que esté presente en varios (Hermisson y Pennings, 2005).

Hay muchos estudios previos dedicados a estudiar el efecto del tamaño poblacional en la adaptación de los virus RNA (Domingo y col., 1999; 2001). En general, los tamaños poblacionales elevados permiten la competencia entre diferentes mutantes, lo cual conduce a la selección positiva de los genomas más eficientes, aumentando así el *fitness* de la población (Martínez y col., 1991; Clarke y col., 1993; Novella y col., 1995 a, b; Escarmís y col., 1998; Arias y col., 2001). Por el contrario, los cuellos de botella poblacionales generan reducciones de la diversidad genética, dando lugar a la fijación rápida de mutaciones, independientemente de su valor selectivo, debido a la disminución de la competencia entre diferentes mutantes. De este modo, el *fitness* de la población se ve disminuido al no estarse seleccionando los mutantes más eficaces y pudiéndose llegar, en casos extremos, a la extinción de la población (Domingo y col., 1996; 1999; Domingo y Holland, 1997; Escarmís y col., 1999; Lázaro y col., 2003; Cases-González y col., 2008). Este mecanismo fue propuesto por primera vez por Müller en 1964 como causa frecuente de extinción en poblaciones pequeñas de organismos asexuales, debido a la acumulación de mutaciones deletéreas (Müller, 1964).

Aunque se ha indicado que los tamaños poblacionales elevados disminuyen la pérdida de mutaciones beneficiosas por deriva génica, cuando la tasa de error es elevada, como ocurre en el caso de los virus RNA, la situación se hace más compleja. Mientras que a baja tasa de error hay suficiente tiempo para la fijación de la primera mutación beneficiosa que aparece en una población antes de que aparezca la siguiente, a tasa de error elevada puede suceder que aparezca una nueva mutación beneficiosa antes de que se haya fijado la primera. En ausencia de recombinación, pueden generarse varias líneas clonales, cada una con una mutación o una combinación de mutaciones diferente, que compiten entre ellas. Como consecuencia, puede producirse la pérdida de mutaciones beneficiosas y retrasarse la llegada a un óptimo adaptativo. Este proceso de competición entre mutaciones beneficiosas se denomina interferencia clonal (Miralles y col., 1999). Fue expuesto inicialmente de forma teórica (Gerrish y Lenski, 1998) y ha sido demostrado después en múltiples estudios llevados a cabo con diferentes sistemas biológicos (Desai y col., 2007; Miller y col., 2011; Strelkova y Lässig, 2012).

Las consideraciones expuestas indican que, aunque en principio, partir de una población más heterogénea podría favorecer y acelerar la adaptación, si esta población contiene ya varios mutantes beneficiosos, la interferencia entre ellos podría tener consecuencias negativas. En ese caso, la reducción del tamaño poblacional podría conducir a la eliminación de alguno de esos mutantes, aumentando así la velocidad de la adaptación.

Que la adaptación ocurra a partir de mutaciones preexistentes o generadas *de novo* también va a tener una gran influencia en el grado de convergencia evolutiva alcanzado entre líneas que evolucionan en paralelo a partir de la misma población ancestral. Si la disponibilidad de mutaciones beneficiosas es elevada y, dado que estas ocurren por azar, probablemente va a tener como consecuencia que diferentes líneas adquieran diferentes mutaciones. Sin embargo, si la adaptación ocurre a partir de la diversidad preexistente, es más probable que diferentes líneas se adapten a través de las mismas mutaciones, dando lugar a un mayor grado de convergencia evolutiva. Este hecho se ha

observado en poblaciones de levaduras (Burke y col., 2014), *Drosophila* (Burke y col., 2010; Kang y col., 2016), *Caenorhabditis elegans* (Chandler, 2014), conejos (Alves y col., 2019) y otros sistemas.

2.6 EL BACTERIÓFAGO Q β

En esta tesis doctoral se ha realizado un estudio de evolución experimental utilizando como sistema el bacteriófago Q β , un virus con genoma de RNA que infecta la bacteria *Escherichia coli*. Este virus posee todas las características que se han mencionado para los virus de RNA (altas tasas de error, estructura en cuasiespecies y elevada velocidad de evolución) y cuenta con las ventajas adicionales de tener un hospedador fácilmente cultivable y poder ser manipulado y propagado de forma sencilla sin requerir estrictas medidas de seguridad.

Desde que el bacteriólogo Frederick Twort (1915) y el microbiólogo Félix d'Herelle (1917) descubrieran la existencia de los bacteriófagos, estos han sido un elemento clave para el estudio de muchos procesos celulares. Los bacteriófagos constituyen las entidades biológicas más abundantes en la Tierra. Regulan la diversidad y la abundancia de las poblaciones bacterianas, lo que los sitúa en la base de la mayoría de los ecosistemas, contribuyendo además a los ciclos biogeoquímicos en la biosfera.

2.6.1 Características generales

El bacteriófago Q β pertenece a los llamados colifagos de ssRNA, que fueron descritos por primera vez por Tim Loeb y Norton Zinder (Loeb y Zinder, 1961). Se clasifica dentro de la familia Leviviridae y el género Allevivirus (Olsthoorn y Van Duin, 2011). Posee una cápsida de simetría icosaédrica T = 3 (Caspar y Klug, 1962) (Fig. 5), que tiene un tamaño de 25 nm de diámetro y que consiste principalmente de 180 copias de la proteína mayoritaria de la cápsida (proteína coat o de cubierta). Estas proteínas se unen entre sí por enlaces disulfuro, dando lugar a pentámeros y hexámeros en relación estequiométrica 12:20, que es consistente con la simetría icosaédrica. En la cápsida hay además de 3 a 5 unidades de la proteína A1 y una unidad de la proteína A2 (Takamatsu y Iso, 1982).

Este bacteriófago tiene un genoma de RNA de cadena sencilla de polaridad positiva, +ssRNA, (como el virus de la polio, el virus de la fiebre aftosa o el SARS-Cov-2) de 4217 nucleótidos que codifican cuatro proteínas (Weissmann y col., 1973; Weissmann, 1974): proteína de cubierta, A1, A2, y la replicasa (Fig. 6). La proteína de cubierta (codificada por los nucleótidos del 1344 al 1742) es la proteína mayoritaria de la cápsida. La proteína A2, también llamada de maduración (codificada por los nucleótidos del 61 al 1320) participa en la unión del fago al pili bacteriano, que actúa como receptor para el virus (Moya y col., 2004), y en la lisis de la bacteria. La proteína A1 (desde el nucleótido 1344 al 2330) (Weiner y Weber 1971; Hofstetter y col., 1974) corresponde a una versión extendida de la proteína de cubierta y, por el momento se desconoce cual es su función.

Por último, la replicasa (nucleótidos del 2352 al 4118) es la proteína encargada de realizar la copia del RNA viral (Moya y col., 2004).

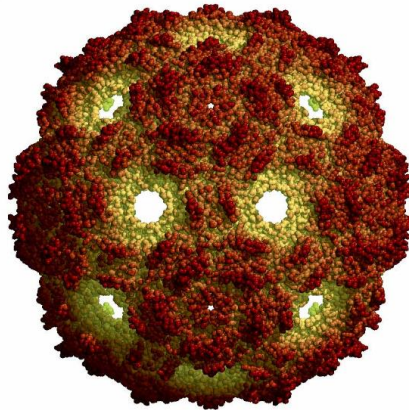


Figura 5: Reconstrucción de la cápside del bacteriófago Qβ.

Imagen de virology.wisc.edu/virusworld

La proteína de la cubierta y la proteína A1 comparten el mismo codón de inicio, aunque esta última solo se sintetiza cuando se produce la incorporación errónea de triptófano en el lugar del codón de terminación de la proteína de cubierta, lo que ocurre con una frecuencia de aproximadamente el 5 % (Weiner y Weber, 1971; Hofstetter y col., 1974). Este fenómeno origina las denominadas proteínas readthrough, y se ha documentado tanto en eucariotas como en procariotas (Kohli y Grosjean, 1981; Valle y Morch, 1988). Este mecanismo de traducción está acoplado, y garantiza que cada genoma produce el número correcto de cada una de estas dos proteínas.

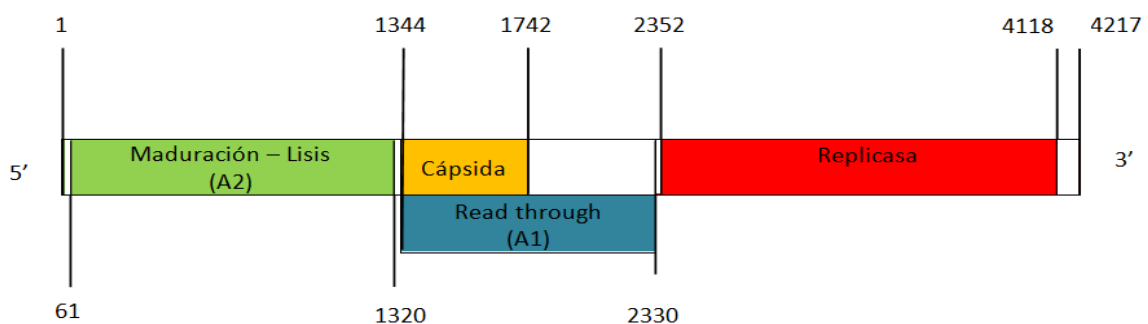


Figura 6: Mapa genómico del bacteriófago Qβ. En él se indican las cuatro proteínas virales, indicando el nucleótido de inicio y fin para cada gen.

La temperatura óptima para la replicación de Qβ es de 37 °C (la temperatura óptima del huésped). Como en otros virus de RNA, la replicación se produce con tasa de error muy elevada, del orden de 1.4×10^{-4} s/n/r (Bradwell y col., 2013), por lo que comete uno o dos errores por cada replicación. Esta tasa de error, sumada al elevado tamaño de sus

poblaciones, supone que estas contienen una gran diversidad genética y, por tanto, evolucionan a gran velocidad. Además, su reducido tamaño genómico permite establecer relaciones entre el genotipo y el fenotipo más fácilmente que en otros sistemas biológicos.

2.6.2 Infección viral

El ciclo infectivo del bacteriófago Q β se divide en tres procesos: i) adsorción y entrada en la célula hospedadora, ii) síntesis de nuevas moléculas de RNA y proteínas virales y iii) lisis celular y liberación de las nuevas partículas virales (Tsukada y col., 2009; García-Villada y Drake, 2013).

i) Adsorción y entrada en la célula hospedadora

El bacteriófago Q β infecta a su hospedador adhiriéndose a los costados del pili bacteriano a través de la proteína de maduración A2 (Fig. 7). Al producirse el contacto la proteína se fragmenta en dos y permite la liberación del material genético dentro de la bacteria. Se cree que la contracción del pili bacteriano hacia el interior celular ayuda a la entrada del material genético viral (van Duin y Tsareva, 2006).

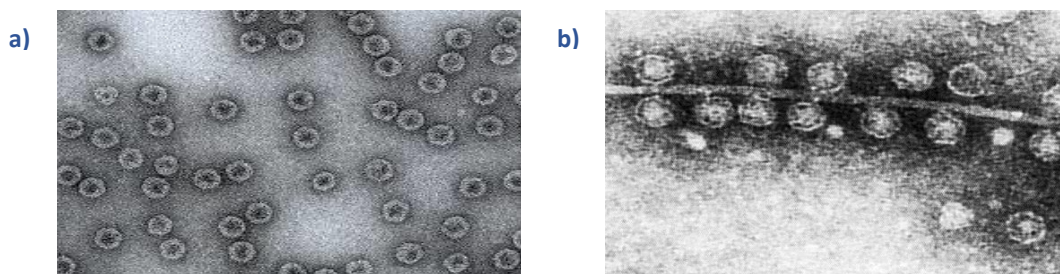


Figura 7: Vistas al microscopio de partículas de Q β . a) partículas libres, b) partículas unidas al F pili de *E. coli*.

ii) Síntesis de nuevas moléculas de RNA y proteínas virales

El fago Q β representa uno de los modelos mejor estudiados de síntesis de RNA viral. La polimerasa que replica las cadenas de RNA es un complejo de cuatro proteínas formado por la interacción de la replicasa del fago con proteínas del hospedador. La subunidad catalítica β (replicasa) está codificada por el fago, mientras que las otras tres subunidades están codificadas por el genoma bacteriano: subunidad α (proteína ribosomal S1), subunidad γ (factor de elongación de la traducción EF-Tu) y subunidad δ (factor de elongación de la traducción EF-Ts) (van Duin y Tsareva, 2006). Cuando la replicasa del fago se ensambla con estas tres proteínas de la bacteria es cuando se forma la polimerasa activa que dará origen a la replicación viral. Esta maquinaria replicativa amplificará el genoma de Q β aproximadamente 10^4 veces en menos de una hora durante la infección (Weissmann, 1974).

Los procesos de replicación y traducción del genoma están altamente regulados, por lo que cuando comienza uno se interrumpe el otro y viceversa. Hay dos posiciones específicas, los llamados sitios S (nucleótidos 1247 a 1346) y M (nucleótidos 2545-2867), que sirven a la replicasa para unirse al RNA. La unión al sitio M le permite aproximarse al extremo 3' del genoma, pudiendo así iniciar la replicación. La unión de la replicasa al sitio S impide la unión de los ribosomas en el lugar donde debería iniciarse la traducción de la proteína de cubierta, impidiendo así que este proceso tenga lugar. La hebra positiva da lugar a la hebra negativa, y ésta es a su vez será el molde de nuevas hebras positivas. Ambas hebras (positiva y negativa) son totalmente complementarias, pero no se alinean como doble hebra debido al alto grado de plegamiento de sus estructuras secundarias y terciarias (Biebricher y col., 1984; Axelrod y col., 1991; Skripkin y Jacobson, 1993). Cuando hay suficiente RNA viral, la proteína de cubierta se une a una región al principio del gen de la replicasa, conocida como el operador e interrumpe la síntesis de esta proteína, lo que tiene como consecuencia el cese de la replicación (Rumnieks y Tras, 2018).

Se ha determinado la estructura cristalina del complejo entre la subunidad β y las dos proteínas huésped a una resolución de 2.5 Å (Kidmose y col., 2010) (Fig. 8). La subunidad β tiene la apariencia de una mano derecha ahuecada conocida de otras estructuras de polimerasas (Ng y col., 2008) y puede ser dividida en tres dominios: la palma, el pulgar y los dedos (Ferrer-Orta y col., 2004). Todas las polimerasas de la familia Leviviridae tienen una estructura similar a la de la subunidad β de Q β .

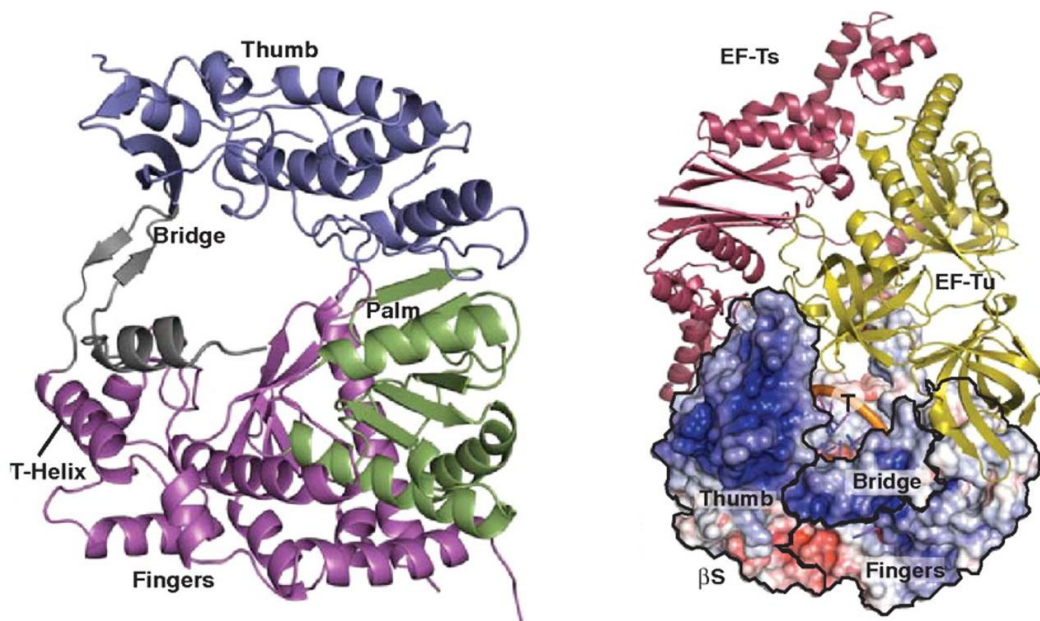


Figura 8: Estructura cristalina del complejo entre la subunidad β y las dos proteínas del huésped a una resolución de 2.5 Å (Kidmose y col., 2010).

Cuando las proteínas virales ya se han sintetizado y se ha copiado el material genético, se produce la formación de las nuevas partículas virales y la encapsidación del RNA.

iii) Lisis celular y liberación de las nuevas partículas virales

Este es un paso crucial para la propagación de la infección de célula a célula. Los nuevos fagos son liberados por medio de la lisis celular. Esta sucede cuando la proteína A2 produce un colapso en la pared celular, al inhibir la acción de la proteína MurA que es la encargada de la síntesis del peptidoglicano (Karnik y Billeter, 1983; Winter y Gold, 1983).

En esta tesis doctoral, el bacteriófago Q β se ha utilizado como modelo experimental para estudiar la influencia de la diversidad genética preexistente y de la pauta de cambio de las presiones selectivas en la adaptación a temperatura elevada. A continuación haremos una breve revisión de los estudios previos sobre la adaptación viral a los cambios de temperatura.

2.7 LA TEMPERATURA COMO PRESIÓN SELECTIVA EN LA ADAPTACION VIRAL

La temperatura es uno de los parámetros ambientales con mayor influencia en prácticamente todos los procesos biológicos. Afecta a la estabilidad de las proteínas (influyendo en sus plegamientos tridimensionales), a la velocidad y precisión de la catálisis enzimática y, posiblemente a la tasa de error de la replicación. No es exagerado decir que la temperatura probablemente es una de las condiciones que más ha influido en la evolución de la vida.

Una observación común en estudios genéticos es que muchas mutaciones que son bien toleradas a la temperatura de crecimiento estándar, sin embargo se comportan como deletéreas a temperaturas solo 5 °C o 10 °C más altas, lo que sugiere que la estabilidad macromolecular es más dependiente de la integridad estructural conforme aumenta la temperatura (Chen y Shakhnovich, 2010). El hecho de que los organismos termófilos acumulen menos mutaciones no sinónimas (con respecto a las sinónimas) estaría de acuerdo con la existencia de mayores restricciones estructurales en sus proteínas (Friedman y col., 2004). El mayor impacto deletéreo de las mutaciones en termófilos podría verse reducido si las tasas de error de la replicación fueran más bajas en estos organismos, algo que se ha sugerido en algunos estudios (Drake, 2009).

La importancia de la temperatura en el *fitness* de las poblaciones, unida al hecho de que este parámetro es fácilmente manipulable en el laboratorio, hace que se hayan realizado numerosos estudios de evolución experimental (muchos de ellos llevados a cabo con virus) en los que se estudia algún aspecto de la adaptación a esta condición.

Uno de los primeros estudios de adaptación viral a temperatura elevada en el que se correlacionaron cambios de *fitness* con cambios genéticos fue realizado en 2001 (Holder y Bull, 2001). En este estudio, utilizando dos bacteriófagos de DNA se demostró que ambos virus fijan primero mutaciones de gran efecto, que serían responsables de los mayores aumentos de *fitness* observados. A partir de este punto, la adaptación puede

continuar, pero de forma más lenta y a través de mutaciones de efecto menor. Las poblaciones de este estudio fueron utilizadas más adelante para explorar la existencia de posibles costes en la *fitness* a distintas temperaturas (Knies y col., 2006). Estos costes pueden resultar de la incapacidad para optimizar simultáneamente la estabilidad de las proteínas y su función, de modo que las enzimas que son estables a altas temperaturas son menos funcionales a temperaturas más bajas, mientras que las enzimas que funcionan mejor a temperaturas bajas son menos estables a altas (Fields, 2001). Aunque los resultados del estudio no fueron concluyentes para identificar costes en *fitness*, sí se identificaron cambios en la forma de las curvas que representan la tasa de crecimiento del virus a distintas temperaturas. Algunos estudios posteriores mostraron resultados dispares, dependiendo del virus y las temperaturas analizadas. Mientras que poblaciones de Q β que se habían adaptado a 43 °C mostraron un coste en *fitness* a 30 °C, pero no a 37 °C (Arribas y col., 2014), en un estudio realizado con el virus de la estomatitis vesicular no se observaron costes en *fitness* a ninguna de las temperaturas ensayadas (Alto y col., 2013).

También se han realizado estudios en los que lo que se analiza no es la adaptación del virus para replicarse a temperaturas no óptimas, sino su capacidad para aumentar su resistencia a temperaturas elevadas en el medio extracelular (Cox y col., 2010). Aunque existen ciertas evidencias de que las tasas de crecimiento altas están asociadas a menor estabilidad extracelular (De Paepe y Taddei, 2006), los resultados de otros estudios no son uniformes como para concluir que de modo general no se pueden optimizar simultáneamente ambos rasgos (Mc Gee y col., 2014; Shingal y col., 2017; Lázaro y col., 2018). Un hallazgo interesante es que las poblaciones virales genéticamente más robustas se adaptan más fácilmente al aumento de la temperatura en el medio extracelular (Mc Bride y col., 2008). Lo mismo funciona a la inversa, es decir, las poblaciones virales que han aumentado su termoestabilidad también son más robustas genéticamente (Domingo-Calap y col., 2010).

Es remarcable que, en los estudios en los que se ha analizado, la adaptación a alta temperatura, tanto cuando el virus está replicando como cuando está en el medio extracelular, ocurre principalmente a través de mutaciones que aumentan la estabilidad de las proteínas o a las interacciones entre ellas (Lee y col., 2011; Miller y col., 2014; Whittington y Rokytá, 2019).

Respecto a nuestro sistema particular de estudio, el bacteriófago Q β , también se han realizado estudios previos sobre su adaptación a temperatura elevada. Los experimentos realizados en nuestro grupo siempre se han llevado a cabo con el virus obtenido tras la expresión del clon infeccioso pBRT7Q β que contiene un cDNA del genoma del virus clonado en pBR322 (Taniguchi y col., 1978; Barrera y col., 1993). Nuestros resultados más importantes son: i) el virus es capaz de adaptarse a 43 °C a través de una variedad de mutaciones que afectan a todos los genes del virus, ii) la adaptación a 43 °C tiene cierto coste en la *fitness* a 30 °C, aunque el virus es capaz de adaptarse a la alternancia de estas dos temperaturas (Arribas y col., 2014), iii) las poblaciones virales que se han propagado en presencia de 5-azacitidina, un mutágeno

que aumenta la tasa de error de la replicación viral, fijan rápidamente un conjunto de mutaciones cuando se propagan a temperatura elevada (Arribas y col., 2016) y iv) la heterogeneidad del espectro de mutantes influye en el valor selectivo de algunas de las mutaciones que se seleccionan a temperatura elevada (Arribas y col., 2018).

En estudios llevados a cabo por otros grupos, también sobre la adaptación de Q β a alta temperatura, se utilizó el virus obtenido tras la expresión del clon infeccioso pACYCQ β (Inomata y col., 2012). El alineamiento de los genomas de ambos virus (los clonados en pACYCQ β y pBRT7Q β) mostró la presencia de dos sustituciones silenciosas en este último (C1257U y C2249U). Estas sustituciones junto con G4A, U192C y C2201U (también silenciosas), fueron un requerimiento necesario para que el virus obtenido tras la expresión de pACYCQ β pudiera ser adaptado a alta temperatura (Kashiwagi y col., 2014). Además, en este caso la temperatura tuvo que aumentarse gradualmente para evitar la extinción del virus. Estudios cuantitativos posteriores mostraron que las poblaciones se adaptaron a través de aumentos en la tasa de adsorción, descensos en el periodo latente y aumentos del número de virus producidos por bacteria infectada (Kashiwagi y col., 2018).

En esta tesis doctoral se plantea utilizar el bacteriófago Q β como sistema experimental y la temperatura como presión selectiva para estudiar, en primer lugar, el efecto de la diversidad preexistente en las poblaciones virales en la adaptación a una nueva presión selectiva. Para ello se obtendrán dos poblaciones del virus que difieren en la extensión de diversidad genética que contienen. Estas dos poblaciones se propagarán a temperatura elevada (43 °C) durante sesenta pases seriados. A lo largo del proceso se determinarán las variaciones de *fitness*, los cambios en la secuencia consenso y los cambios en el espectro de mutantes estimados mediante secuenciación masiva.

En segundo lugar se analizará el efecto de la velocidad de cambio de las presiones selectivas en la adaptación. Para ello la misma población ancestral se propagará a temperaturas más altas de la óptima siguiendo diferentes patrones de cambio. Las poblaciones serán caracterizadas a lo largo del proceso evolutivo a nivel genotípico y fenotípico, del mismo modo descrito anteriormente.

3. OBJETIVOS

1. Estudio de la influencia de la diversidad genética intrapoblacional en la dinámica de ganancia de *fitness* y en las vías adaptativas seguidas durante la adaptación del bacteriófago Q β a temperatura más alta que la óptima.
2. Estudio de la influencia de la velocidad de cambio en las presiones selectivas en la dinámica de ganancia de *fitness* y en las vías adaptativas seguidas durante la adaptación del bacteriófago Q β a temperatura más alta que la óptima.
3. Comparación de las variaciones en el *fitness* y las vías evolutivas seguidas por el bacteriófago Q β cuando se propaga de forma constante o alternante a temperaturas mayores y menores que la óptima.
4. Caracterización mediante NGS de los cambios que se producen en la diversidad genética intrapoblacional y en la estructura interna de las poblaciones a lo largo de la adaptación a las variaciones de temperatura en el bacteriófago Q β .
5. Efecto en el *fitness* a distintas temperaturas de las mutaciones adquiridas por el bacteriófago Q β durante su adaptación a temperatura más alta que la óptima.

4. MATERIALES Y MÉTODOS

4.1 MATERIALES

4.1.1 Medios de cultivo y crecimiento

Medio líquido Nutrient Broth (NB)

8 g/l Nutrient Broth (Merck)

5 g/l NaCl (Merck)

Agar semisólido

8 g/l Nutrient Broth (Merck)

5 g/l NaCl (Merck)

7 g/l Select Agar (Invitrogen)

Medio LB líquido

20 g/l LB Broth Base (Lennox L Broth Base) (Invitrogen)

Medio LB Agar

32 g/l LB Agar-Lennox L Agar (Invitrogen)

4.1.2 Disoluciones y tampones

Tampón del fago

1 g/l gelatina (Merck)

50 ml/l Tris-HCL 1 M pH 7.5 (Gibco)

10 ml/l MgCl₂ 1 M (Merck)

Tampón para electroforesis (TAE 1x)

100 ml/l tampón TAE 10x (composición Tris 40 mM, acetato 40 mM y 1 mM EDTA, pH 8.3) (Gibco)

Agarosa para electroforesis

10 g/l Agarosa, grado electroforesis (Invitrogen), diluida en TAE 1x

4.2 MÉTODOS EXPERIMENTALES, BIOINFORMÁTICOS Y TEÓRICOS

4.2.1 Virus y bacterias. Procedimientos generales de infección

Obtención del clon viral Q β _{WT}

El plásmido pBRT7Q β , que contiene el cDNA del bacteriófago Q β clonado en el plásmido pBR322 (Taniguchi y col., 1978; Inomata y col., 2012) se usó para transformar *Escherichia coli* DH5- α (Invitrogen), la cual puede expresar los genes virales y realizar el ensamblaje de sus partículas, pero no puede ser infectada ya que carece del F pili. El sobrenadante de un cultivo ON, obtenido a partir de una colonia transformada, se utilizó para infectar *E. coli* Hfr (Hayes) en agar semisólido, a una multiplicidad de infección (mdi) que permitió la generación de placas líticas bien separadas. En estas condiciones, hay una alta probabilidad de que cada placa lítica sea el resultado de la replicación de un solo virus en un número reducido de generaciones y, por tanto, puede considerarse un clon biológico.

La progenie del virus contenida en una placa lítica elegida al azar se extrajo recogiendo el agar ocupado por la placa e incubándolo en 1 ml de tampón del fago con 50 μ l de cloroformo, durante 1 h a 28 °C y 850 rpm en termobloque. A continuación, se centrifugó durante 15 min a 12000 rpm y se recogió el sobrenadante conteniendo las partículas virales. La población viral clonal así preparada fue secuenciada (ver apartado 4.2.5. de Materiales y Métodos), comprobándose que no tenía ninguna mutación respecto a la secuencia del cDNA clonado en pBR322. Para nuestros estudios, este clon fue considerado un análogo del virus salvaje, se denominó Q β wild type (Q β _{WT}) y fue el origen de todas las poblaciones generadas en esta tesis.

Infección en medio líquido y determinación del título viral

Para el mantenimiento y la propagación del bacteriófago Q β en medio líquido se utilizaron cultivos de *E. coli* Hfr en fase exponencial (DO₅₅₀ entre 0.6 y 0.8). Para su obtención, se prepararon diluciones 1:20 de cultivos en fase estacionaria (obtenidos a partir de colonias aisladas incubadas ON en NB a 37 °C y 250 rpm) y se incubaron durante 2 horas a 37 °C y 110 rpm.

Las infecciones virales se realizaron a la mdi indicada en cada experimento, utilizando bacterias de fase exponencial, que se diluyeron en proporción 1:1 hasta completar un volumen de 1 ml en medio NB. Inicialmente estos cultivos contienen del orden de 10⁸ bacterias. Tras 2 horas de incubación a 37 °C y 250 rpm, los cultivos fueron tratados con 1/20 volúmenes de cloroformo, se incubaron en un termobloque durante 15 min a 37 °C y 600 rpm y, a continuación se centrifugaron durante 10 min a 12000 rpm. Finalmente se recogieron los sobrenadantes, se añadió un volumen 1/40 de cloroformo y se guardaron a 4 °C, si iban a ser usados en un plazo máximo de 4 semanas, o a - 80 °C para su conservación a largo plazo.

Los títulos virales se determinaron mediante ensayo en placa, para lo cual se mezclaron 300 μ l de bacterias *E. coli* Hfr en fase exponencial, 100 μ l de la suspensión de fago y 4.5 ml de agar semisólido y se vertieron sobre una placa de Petri, que contenía una capa inferior de medio LB agar. La concentración de fagos se ajustó, realizando las oportunas diluciones con tampón del fago, para obtener entre 20 y 40 placas de lisis. Los títulos se expresaron como el número de unidades formadoras de placas (ufp) por ml de la suspensión de fagos.

4.2.2 Propagación del virus Q β _{WT}

Obtención de las poblaciones ancestrales

El virus Q β _{WT} (apartado 4.2.1.) se utilizó para infectar un cultivo de *E. coli* Hfr en fase exponencial, usando una mdi de 0.1, en un volumen final de 1 ml, como se ha descrito en el apartado anterior. Después de 2 horas de incubación a 37 °C y del tratamiento del cultivo con cloroformo, se recogió el sobrenadante conteniendo las partículas virales, y una fracción del mismo se utilizó para infectar un nuevo cultivo de *E. coli*, manteniendo la mdi alrededor de 0.1. Este proceso se repitió 25 veces, correspondiendo cada una de ellas a una transferencia o pase seriado (Fig. 9).

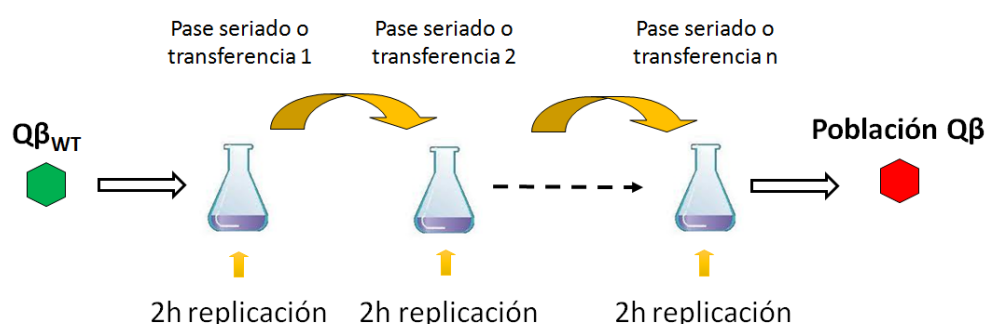


Figura 9: Esquema del procedimiento experimental para la propagación del bacteriófago Q β mediante el sistema de pases seriados.

Las poblaciones obtenidas tras 2 pases seriados (Q β _{P2}) y 25 pases seriados (Q β _{P25}) fueron el origen (poblaciones ancestrales) de todas las líneas evolutivas analizadas en esta tesis doctoral.

Líneas evolutivas obtenidas a partir de las poblaciones ancestrales

El ancestro $Q\beta_{P25}$ se transfirió durante 60 pases seriados a través de tres patrones de cambio de temperatura diferentes (Fig. 10); uno en el que la temperatura se aumentó directamente a 43 °C, manteniéndose constante en ese valor durante toda la serie de pases (C43), y otros dos en los que la temperatura de 43 °C se alcanzó de forma gradual (G1 y G2). El patrón G1 supuso la propagación del virus durante 30 pases a 40 °C y otros 30 pases a 43 °C, mientras que el patrón G2 implicó aumentos de un grado cada 10 transferencias, comenzando en 38 °C los 10 primeros pases y alcanzando los 43 °C en las transferencias número 51 al 60. Todas las propagaciones se realizaron en triplicado resultando los conjuntos de líneas evolutivas $C43_{P25}$, $G1_{P25}$ y $G2_{P25}$ (Fig. 10). El subíndice P25 indica que el origen de estas poblaciones fue la población $Q\beta_{P25}$.

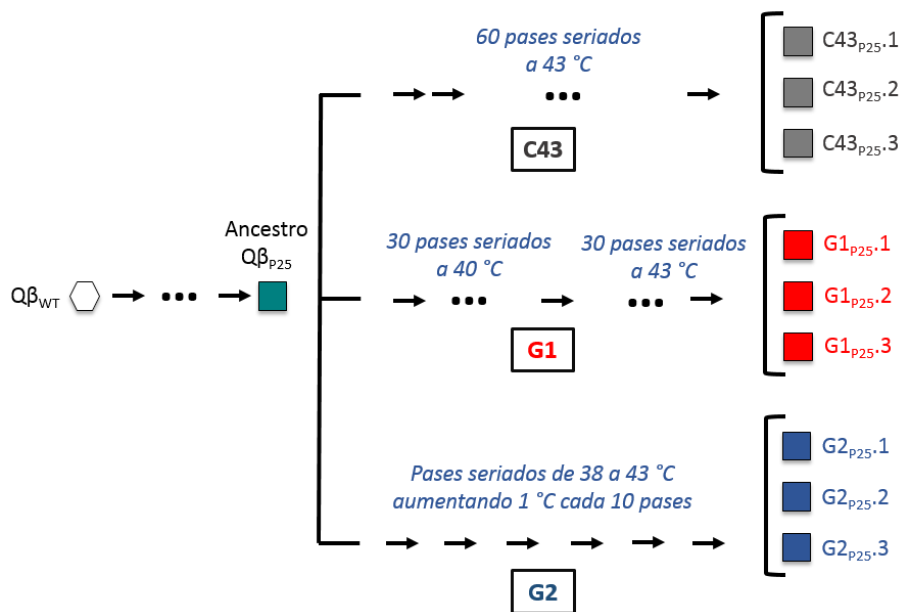


Figura 10: Esquema mostrando la evolución de la población ancestral $Q\beta_{P25}$ a través de distintos patrones de cambio de temperatura. La propagación de forma constante a 43 °C dio lugar a las líneas evolutivas $C43_{P25}$, mientras que los aumentos de temperatura graduales dieron lugar a las líneas evolutivas $G1_{P25}$ y $G2_{P25}$. En todos los casos se realizaron tres réplicas, obteniéndose así las poblaciones denotadas por 1, 2 y 3.

El ancestro $Q\beta_{P2}$ se transfirió durante 60 pases seriados en los que la temperatura se mantuvo constante a 30 °C, 33 °C, 37 °C, 40 °C o 43 °C (patrones de cambio C30, C33, C37, C40 y C43; Fig. 11). En el caso del patrón de cambio alternante (A) se alternaron las temperaturas de 30 °C y 43 °C cada pase (Fig. 11). Como en el caso anterior, todas las propagaciones del virus se realizaron en triplicado, resultando los conjuntos de líneas

evolutivas C30_{P2}, C33_{P2}, C37_{P2}, C40_{P2}, C43_{P2} y A_{P2}. El subíndice P2 indica que el origen de estas poblaciones fue la población Q β _{P2}.

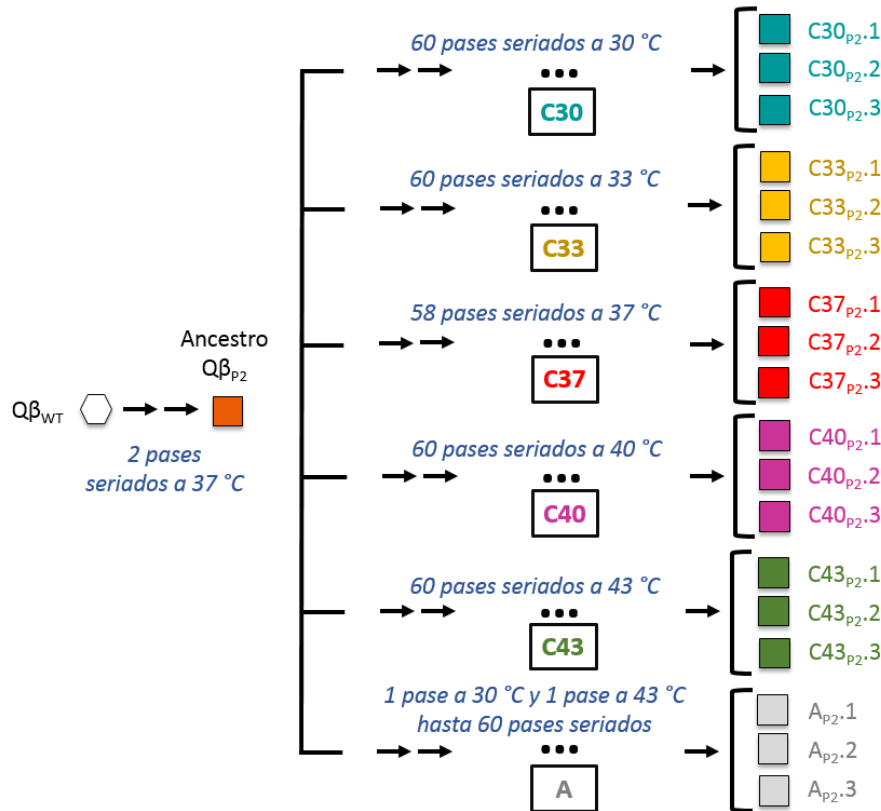


Figura 11: Esquema mostrando la evolución de la población ancestral Q β _{P2} a través de distintos patrones de cambio de temperatura. Para cada uno de los patrones se realizaron tres réplicas, obteniéndose así las poblaciones denotadas por 1, 2 y 3. Cabe señalar que la población ancestral Q β _{P25} es la réplica 3 de la población C37_{P2} en el pase 23 de su evolución.

4.2.3 Ensayos de determinación de *fitness*

El *fitness* de las poblaciones virales se estimó como el valor de su tasa de crecimiento, o capacidad replicativa, en las condiciones utilizadas en el ensayo.

Los ensayos se realizaron inoculando por duplicado cultivos líquidos que contenían 10⁸ bacterias en fase exponencial con 10⁴ ufp en un volumen final de 1 ml. Después de dos horas de incubación a la temperatura indicada, se recogieron los sobrenadantes como se ha descrito anteriormente y se titularon para estimar el número de ufp obtenidas.

El *fitness* absoluto se expresó como el número de duplicaciones en 2 horas, y se calculó como $\log_2 [(N_f - N_i) / N_i]$, donde N_i es el número de ufp iniciales y N_f son las ufp finales obtenidas tras la incubación. Los valores de *fitness* relativo se obtuvieron dividiendo los

valores absolutos por los valores de *fitness* de los ancestros correspondientes, determinados en el mismo ensayo.

En experimentos anteriores llevados a cabo por nuestro grupo se demostró que, cuando las infecciones se llevan a cabo en estas condiciones, la concentración de bacterias se mantiene en exceso sobre la de fagos durante las dos horas de incubación.

4.2.4 Obtención de mutantes únicos mediante mutagénesis dirigida

El proceso de mutagénesis dirigida se llevó a cabo con el kit Quick-Directed Mutagenesis Kit (Agilent Technologies) utilizando parejas de cebadores complementarios específicos que contienen la sustitución que se desea introducir en el genoma viral (Tabla 1). El producto resultante se digirió con la enzima DpnI y se utilizó para transformar por choque térmico células competentes DH5- α (Invitrogen) (Woodcock y col., 1988). Las colonias transformadas se crecieron durante toda la noche a 37 °C con 250 rpm de agitación. Los cultivos resultantes se centrifugaron y los sobrenadantes se plaquearon junto con bacterias *E. coli* Hfr en agar semisólido para la obtención de placas de lisis, las cuales corresponden a clones biológicos del virus que llevan incorporada la mutación deseada. Cada uno de los experimentos llevados a cabo con estos mutantes fue iniciado con una única placa lítica, cuya secuencia fue previamente analizada para comprobar la presencia de la mutación deseada.

Sustitución	Parejas de cebadores
A1088G	5' GTCAGCTATATCATAATATCGGTATTGTAGACGGCTTTGACAG 3'
	5' CTGTCAAAGCCGTCTACAATACCGATATTATGATATAGCTGAC 3'
G1312A	5' TATCTTTTATTAACCCAACGCATAAAGCGTTGAACTTTGGGTC 3'
	5' GACCCAAAGTTTCAACGCTTTATGCGTTGGGTTAATAAAAAGATA 3'
C1806T	5' CCGGTTATTCCGGATCCATCGATTGATCCGCC 3'
	5' GGCGGATCAATCGATGGATCCGAATAACCGG 3'
A1930G	5' TAATGCTGTTGAACTCCGCCTCGGAATTTGATG 3'
	5' CATCAAATTCGCGAGGCCGGAGTTCAACAGCATT 3'
T2776C	5' GCTAATAGGAGATGCTCCGTCCGTTGAGGG 3'
	5' CCCTCAACGGACGGAGCATCTCTATTAGC 3'
T3311G	5' GTTGTTACCTACGAGAAGATGCTTCTATGGGTAACGG 3'
	5' CCGTTACCATAGAAGACATCTTCTCGTAGGTAACAAC 3'
C3545T	5' TTCAGAGAGTCGTGCGGTAAAGCACTACTATTCTGG 3'
	5' CCAGAATAGTAGTGCTTACCGCAGCACTCTCTGAA 3'

Tabla 1: Secuencias genéticas de las parejas de cebadores utilizados para obtener los mutantes de Q β conteniendo las mutaciones indicadas. En cada caso se utilizaron un par de cebadores complementarios conteniendo la mutación deseada (indicada en rojo).

4.2.5 Determinación de las secuencias consenso de las poblaciones de Q β

Extracción de RNA viral

Los sobrenadantes virales se trataron con 1/10 vol de SDS al 10 % (Gibco-Invitrogen), 1/10 vol β -mercaptoetanol 14.3 M (Sigma) y 1 μ l de RNAsina (Promega) durante 20 min a 25 °C y 650 rpm en termobloque. Después de enfriar 5 min en hielo, el RNA se mezcló con 1 volumen de fenol (Invitrogen) saturado con agua fría y se centrifugó durante 5 min a 4 °C y 12000 rpm. La fase acuosa (superior) se precipitó con 1/10 vol de NaAc 3 M (pH 5.5) y 2 vol de etanol absoluto (Merck) durante 30 min a 4 °C. A continuación, se centrifugó durante 10 min a 12000 rpm y 4 °C, y se eliminó el sobrenadante resultante.

Tras lavar con etanol (70 %), el RNA se resuspendió en 20 μ l de agua libre de RNAasas (Ambion) y se almacenó a - 80 °C.

Obtención de cDNA y amplificación por PCR

El RNA viral se utilizó para la síntesis de cDNA. El proceso de retrotranscripción se llevó a cabo utilizando la enzima AMV Transcriptasa Reversa (Promega) obtenida de partículas virales del virus de la mieloblastosis aviar. El cDNA obtenido se amplificó mediante el sistema Expand High Fidelity RT-PCR System (Roche).

El proceso de PCR consiste en la amplificación de una secuencia de nucleótidos comprendida entre dos cebadores o *primers* seleccionados. Estos cebadores se alinean con secuencias específicas de la molécula de DNA y permiten la amplificación de la región entre ambos. El proceso consistió en un programa de 40 ciclos de desnaturalización a 94 °C, alineamiento a una temperatura específica para cada pareja de cebadores utilizados (Tabla 2) y una fase de extensión a una temperatura de 68 °C.

Región amplificada	Pareja de cebadores	Secuencia genética del cebador	Temperatura de alineamiento (°C)
150-1497	P6 (D)	5' CGAATCTTCCGACACGCATCC 3'	56.8
	P7 (R)	5' AAACGGTAACACGCTTCTCCAG 3'	
1407-2817	P8 (D)	5' CTCAATCCGCGTGGGGTAAATCC 3'	59.4
	P9 (R)	5' CAGAAAATCGGCAGTGACGCAACA 3'	
2154-4195	PD5 (D)	5' GTGCCATACCGTTTGACT 3'	55.2
	PR12 (R)	5' GATCCCCCTCTCACTCGT 3'	

Tabla 2: Cebadores utilizados en las reacciones de RT-PCR. Se indican las posiciones nucleotídicas iniciales y finales de las regiones genómicas amplificadas por cada par de cebadores, así como sus correspondientes temperaturas de alineamiento. D y R indican directo y reverso, respectivamente.

Los productos de la amplificación se analizaron mediante electroforesis en gel de agarosa 1 % (con TAE 1x) y fueron visualizados con un transiluminador de UV. Para los geles de agarosa se utilizó el agente intercalante SYBR™ Safe DNA Gel Stain (Invitrogen), el marcador DirectLoad™ 1 kb DNA Ladder (Sigma) y el buffer de carga GelPilot DNA LoadingDye 5x (QIAGEN).

La purificación de los productos de PCR se realizó con el kit de purificación por columna QIAquick PCR Purification Kit (QIAGEN) y la concentración del DNA resultante se cuantificó espectroscópicamente (Nanodrop ND-1000, Applied Biosystems).

Secuenciación del cDNA y análisis de las secuencias

Una vez purificados, los productos de PCR se sometieron a secuenciación estándar por el método de Sanger, en la Unidad de secuenciación del Centro de Astrobiología, utilizando un secuenciador automatizado Abi 3730 XL con Big Dye Chemistry v3.1 (Applied Biosystems; Perkin Elmer), o se enviaron a secuenciar a la empresa Macrogen. Los oligonucleótidos utilizados como cebadores fueron los mostrados en la Tabla 3, los cuales nos permitieron determinar las secuencias comprendidas entre los nucleótidos 280 y 4150.

Región amplificada	Cebador	Secuencia genética
150-1497 (P6/P7)	Psc1	5' TGTGAACGCCACCTTGATGAT 3'
	Psc2	5' TGGCAGAACCGTCATGATAAGATT 3'
1407-2817 (P8/P9)	Psc3	5' CGTTGCCTCGCTTCACA 3'
	PD3	5' CCGATGGCGTGATAGTTG 3'
2154-4195 (PD5/PR12)	PD4	5' CGAAGCTGTTGCCTGGGAGAAGTT 3'
	PD10	5' CCAACTGGGTATCGGTGGCATTCTA 3'
	Psec6	5' CCTGATGGTTACGGTGATGG 3'

Tabla 3: Cebadores utilizados para la determinación de las secuencias consenso. Se muestran las regiones amplificadas por cada pareja de primers (Tabla 2) y los oligonucleótidos utilizados para secuenciarlas.

Las secuencias se ensamblaron con el programa SeqManPro (DNASTAR Lasergene 12 Core Suite) y se alinearon con la del virus Q β _{WT} utilizando el programa Clustal X2.1 (modo de alineamiento múltiple y parámetros predeterminados). Las mutaciones respecto a la secuencia del virus Q β _{WT} fueron identificadas utilizando el editor de alineamientos de secuencias BioEdit v.7.0.5. Para hacer un análisis más completo, las mutaciones, ya sean las fijadas o las polimórficas, que se detectaron en cada línea evolutiva, se inspeccionaron visualmente en los cromatogramas del resto de líneas para evaluar su posible presencia como polimorfismos no reconocido por los programas de análisis de secuencias. Estos cromatogramas fueron visualizados con ayuda del software Chromas v2.6.4.

4.2.6 Procedimientos y análisis de secuenciación masiva

Preparación de librerías y secuenciación

Para el análisis de las poblaciones virales mediante las tecnologías de Next Generation Sequencing (NGS), o secuenciación masiva, se utilizaron los servicios de la Unidad Genómica del Parque Científico de Madrid.

El RNA viral, purificado con un kit de extracción de alta calidad (Mini Kit de ARN Viral QIAamp; QIAGEN), fue utilizado para generar tres amplicones de 312, 393 y 317 pb (Tabla 4) mediante RT-PCRs llevadas a cabo con enzimas de alta fidelidad para minimizar la introducción de errores. Las enzimas utilizadas fueron SuperScript II Reverse Transcriptase (Invitrogen) para la RT y Q5 Hot Start High fidelity (BioLabs) para la PCR.

Las librerías se secuenciaron en dos proyectos de secuenciación diferentes (Sec1 y Sec2) realizando dobles lecturas (paired-end) con el sistema Illumina Miseq (2x250 bp) (<https://emea.illumina.com/>). Se obtuvieron 10.5 millones de lecturas totales distribuidas de forma homogénea entre todas las muestras secuenciadas en cada carrera. Los cebadores utilizados fueron los mostrados en la Tabla 4.

Amplicón 1 (1060-1331)*		
Cebador directo	Pos	Tm (°C)
GAATGTTGGTGACATACTTGCT	1037	59.71
Cebador reverso	Pos	Tm (°C)
TTGCCATGATCAAATTGACC	1349	57.44
Longitud del amplicón	312	

Amplicón 2 (2145-2473)*		
Cebador directo	Pos	Tm (°C)
GGTGCTTCGGTAACATTGAG	2123	60.31
Cebador reverso	Pos	Tm (°C)
TGAATGAAATACACTCAGCCTCAG	2516	61.42
Longitud del amplicón	393	

Amplicón 3 (3329-3603)*		
Cebador directo	Pos	Tm (°C)
AGATTTCTTCTATGGGTAACGG	3306	58.10
Cebador reverso	Pos	Tm (°C)
CCAGTATTAAATCGGCAGGAC	3623	59.29
Longitud del amplicón	317	

Tabla 4: Cebadores utilizados para generar los amplicones secuenciados mediante NGS. En cada tabla se indican las secuencias de nucleótidos de cada una de las parejas de cebadores utilizados en las reacciones de RT-PCR, así como su temperatura de melting (Tm) y la posición correspondiente al nucleótido 5'. También se muestra la longitud del amplicón y, en el paréntesis marcado con *, el intervalo de lectura una vez eliminadas las secuencias correspondientes a los cebadores.

Las poblaciones secuenciadas mediante NGS se indican en la Tabla 5.

Población	Pase	Amplícón
Q β _{P2}	2	1, 2 y 3
Q β _{P25}	25	1, 2 y 3
C43 _{P2.3}	1, 10 y 60	1, 2 y 3
C43 _{P25.1}	1, 10 y 60	1, 2 y 3
G1 _{P25.2}	30, 31, 40 y 60	1, 2 y 3
G2 _{P25.1}	50, 51 y 60	1, 2 y 3
A _{P2.2}	60	1, 2 y 3
C30 _{P2.1}	60	1, 2 y 3
C33 _{P2.1}	60	1
C37 _{P2.3}	60	1
C40 _{P2.1}	60	1

Tabla 5: Poblaciones ancestrales y líneas evolutivas secuenciadas mediante NGS. En la tabla se indican los pases y amplicones secuenciados en cada caso.

Procesado de los datos de NGS

Los datos proporcionados por la unidad de secuenciación fueron procesados con *Galaxy*, plataforma de código abierto para la investigación biomédica e intensiva de datos (www.usegalaxy.org) siguiendo el esquema que se muestra en la Fig. 12.

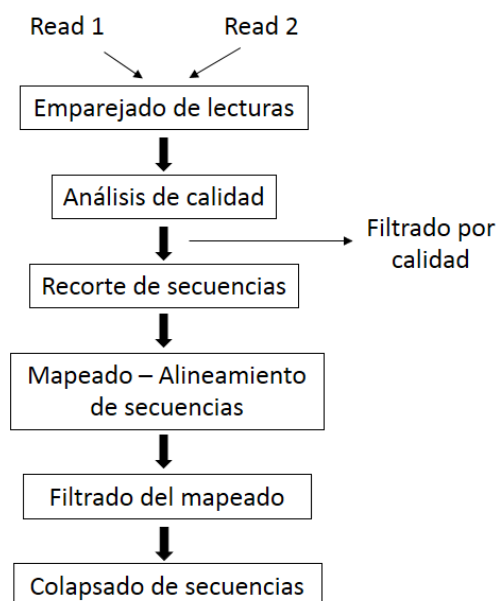


Figura 12: Esquema de flujo del análisis bioinformático.

El primer paso en este análisis fue hacer el emparejado de las secuencias de doble lectura (Magoc y Salzberg, 2011) y, tras él, un estudio de calidad (Babraham Bioinformatics - FastQC; www.illumina.com). Todos los datos poseían una calidad media por base muy buena, con $Q > 30$ (una probabilidad de error de 1 base entre 1000), es decir, una exactitud de asignación de base de un 99.9 %.

Después se eliminaron los primers de secuenciación, recortando las secuencias desde el archivo *fastq*. A continuación, se procedió al mapeado con *BWA*, algoritmo de mapeo contra el genoma de referencia para alineamiento de lecturas de secuencias, implementado en *Galaxy* (Li y Durbin, 2009; 2010; Li, 2013). En todos los casos, el genoma de referencia fue el del virus Q β _{WT}.

Los resultados de los alineamientos, fueron filtrados por posición de alineamiento (considerando que las lecturas mapearan en la región secuenciada correcta) y por cadena CIGAR (que el alineamiento cubriera la región completa del amplicón y que no contuviera ni deleciones ni inserciones).

Debido a la buena calidad y al procedimiento de doble lectura, no fueron necesarios procesos de filtrado por calidad ni recortar el tamaño de las secuencias amplificadas por la común pérdida de esta en los extremos de las lecturas (Barnett y col., 2011; Barnett, 2021).

Para cada una de las muestras y para cada uno de los amplicones correspondientes, se obtuvo como producto de la secuenciación masiva descrita en el apartado anterior el número total de lecturas, N_t , que se muestra en la Tabla 6. El valor de N_t es variable, pero en todos los casos está por encima de 200000. En el segundo proyecto de secuenciación, Sec 2, debido a un error técnico, el número de lecturas de los amplicones 2 y 3 (marcadas con un * en la Tabla 6) fue muy superior al valor medio obtenido en el proyecto Sec1. En esos casos, para tener una toma de datos homogénea, se escogieron aleatoriamente 300000 lecturas de cada muestra, número que entraba dentro del promedio de lecturas que se ha obtenido en los otros casos (327963 ± 169284).

Tras el procesado bioinformático de los datos, el número total de lecturas N_t , se redujo a un valor N_0 . La reducción del número de lecturas se debe a la eliminación de secuencias sin correspondencia en doble lectura, lecturas no completas y/o que mapean mal con la secuencia de Q β _{WT}. Al realizar este procesado bioinformático, las lecturas resultantes proporcionan una cobertura de bases en la secuencia, que en este caso fue un valor homogéneo, ya que hay el mismo número de lecturas para todas las bases que componen la región amplificada en cada una de las muestras en estudio.

El conjunto de lecturas N_0 contiene cierto número de secuencias o haplotipos diferentes (indicados como *Hpl* en la Tabla 6). Estos fueron identificados mediante un colapsado de secuencias, que se llevó a cabo con una de las funciones implementadas en la plataforma *Galaxy* llamada *Collapse*, que contrae secuencias idénticas de un archivo *fasta* en una sola secuencia y cuenta el número de apariciones de esta misma.

	AMPLICIÓN 1				
	Nt	Procesado bioinformático		Umbral 0.05 %	
		N ₀	Hpl	N	Hpl
Q β _{P2}	324304	309124	13688	266858	30
Q β _{P25} (PCR1-Sec1)	286318	274466	14710	235849	84
Q β _{P25} (PCR1-Sec2)	298071	281563	11946	244723	83
Q β _{P25} (PCR2-Sec2)	291044	271465	11193	236742	83
C43 _{P2.3} (P1)	322987	308102	10719	270956	28
C43 _{P2.3} (P10)	294266	281283	13315	239054	35
C43 _{P2.3} (P60)	336153	320518	17844	265606	64
C43 _{P25.1} (P1)	274972	262584	12485	228175	78
C43 _{P25.1} (P10)	295241	283436	13610	242624	85
C43 _{P25.1} (P60)	285328	270487	15333	220119	90
G1 _{P25.2} (P30)	297436	285010	17873	237588	76
G1 _{P25.2} (P31)	260718	250194	17327	207149	73
G1 _{P25.2} (P40)	316778	300060	19445	244983	45
G1 _{P25.2} (P60)	254703	243365	15496	201042	60
G2 _{P25.1} (P50)	306619	293252	20911	237433	62
G2 _{P25.1} (P51)	338250	323463	20606	263451	57
G2 _{P25.1} (P60)	343406	326616	16855	276971	70
A _{P2.2} (P60)	251849	235634	9910	206263	74
C30 _{P2.1} (P60)	238011	223180	10919	188441	76
C33 _{P2.1} (P60)	234385	220602	12776	184375	98
C37 _{P2.3} (P60)	259270	243768	13443	200913	71
C40 _{P2.1} (P60)	218248	205894	11204	171155	76

	AMPLICIÓN 2				
	Nt	Procesado bioinformático		Umbral 0.05 %	
		N ₀	Hpl	N	Hpl
Q β _{P2}	379763	358986	26127	294855	49
Q β _{P25} (PCR1-Sec1)	380690	359406	31345	279452	71
Q β _{P25} (PCR1-Sec2)	900151*	273617	29143	206416	81
Q β _{P25} (PCR2-Sec2)	863790*	271929	29398	203939	77
C43 _{P2.3} (P1)	361672	341163	29652	274730	53
C43 _{P2.3} (P10)	313788	295677	23892	236250	52
C43 _{P2.3} (P60)	351465	329987	32969	252206	78
C43 _{P25.1} (P1)	365297	347057	30977	281442	112
C43 _{P25.1} (P10)	375212	356820	42967	263024	76
C43 _{P25.1} (P60)	970844*	259531	25712	203576	95
G1 _{P25.2} (P30)	469519	444800	45532	330443	64
G1 _{P25.2} (P31)	391401	370458	49218	261548	75
G1 _{P25.2} (P40)	444601	420411	42935	310689	55
G1 _{P25.2} (P60)	411995	388514	43062	289525	85
G2 _{P25.1} (P50)	375228	356918	31877	282673	72
G2 _{P25.1} (P51)	352830	333152	47036	242113	77
G2 _{P25.1} (P60)	323573	307556	28974	242798	71
A _{P2.2} (P60)	862686*	235303	23075	189954	49
C30 _{P2.1} (P60)	783743*	281487	24410	222745	77

	AMPLICÓN 3				
	<i>Nt</i>	Procesado bioinformático		Umbral 0.05 %	
		<i>N₀</i>	<i>Hpl</i>	<i>N</i>	<i>Hpl</i>
Qβ_{P2}	277476	262456	15078	223334	39
Qβ_{P25}(PCR1-Sec1)	439189	408717	20728	351353	75
Qβ_{P25}(PCR1-Sec2)	985123*	274311	11176	240322	81
Qβ_{P25}(PCR2-Sec2)	890829*	270931	11847	236158	80
C43_{P2.3} (P1)	305124	288236	15173	247320	40
C43_{P2.3} (P10)	403218	378041	20695	319696	35
C43_{P2.3} (P60)	389776	367215	25930	300033	58
C43_{P25.1} (P1)	337665	304935	15879	264974	69
C43_{P25.1} (P10)	294201	276028	20515	229209	87
C43_{P25.1} (P60)	739194*	278308	14517	236482	85
G1_{P25.2} (P30)	278402	261241	18328	212969	52
G1_{P25.2} (P31)	359868	337800	24447	275050	54
G1_{P25.2} (P40)	306213	288273	18958	238890	49
G1_{P25.2} (P60)	306554	289177	20082	242864	70
G2_{P25.1} (P50)	289455	272288	18995	228133	70
G2_{P25.1} (P51)	420610	394647	24232	331820	64
G2_{P25.1} (P60)	365297	343924	21567	289698	65
A_{P2.2} (P60)	991679*	276217	11385	242689	62
C30_{P2.1} (P60)	791802*	280568	12282	245275	74

Tabla 6: Número de lecturas y haplotipos obtenidos tras la secuenciación masiva de cada una de las poblaciones en estudio. En cada una de las tablas se indica el número de lecturas totales (*Nt*), número de lecturas obtenidas tras el procesamiento bioinformático (*N₀*), las correspondientes a las secuencias representadas en más del 0.05 % de las lecturas contenidas en *N₀* (*N*) y el número de haplotipos (*Hpl*). La notación PCR1-Sec1, PCR1-Sec2, y PCR2-Sec2 indica diferentes combinaciones de reacciones de PCR y proyectos de secuenciación que se realizaron con la población Qβ_{P25} y cuya descripción detallada se realiza en el Capítulo 1 de Resultados. Al lado de cada línea evolutiva aparece también indicado el número de pase secuenciado (entre paréntesis).

4.2.7 Estudio de la diversidad de las poblaciones virales

Para los análisis de diversidad, se utilizaron las secuencias cuya representación en el número total de lecturas, tras el procesamiento bioinformático, fuera igual o superior al 0.05 % (Tabla 6). El número de lecturas correspondientes a esas secuencias se indica como *N*. Una justificación detallada de la elección del umbral del 0.05 % se discutirá en el Capítulo 1.

Para el estudio de la diversidad genética contenida en las poblaciones virales analizadas por secuenciación masiva se determinaron varios índices, propios del campo de la ecología, que son también de común aplicación en virología. Ya que ningún índice de diversidad puede capturar totalmente la complejidad de una población viral (Gregori y col., 2016), se realizó un análisis multivariante, examinando el comportamiento de diferentes índices para un mismo conjunto de datos.

Los índices de diversidad utilizados se basan en la información proporcionada por la secuencia y distribución de los haplotipos contenidos en cada población. Pueden ser clasificados en:

- **Basados en incidencia**: Reflejan la incidencia de los elementos de diversidad considerados sin tener en cuenta sus frecuencias. Los que se han analizado en la presente tesis son el número de sitios polimórficos (M), el número de mutaciones (nM), el número de haplotipos (Hpl) y la frecuencia de mutación mínima ($Fmín$).
- **Basados en abundancia**: Consideran tanto las entidades observadas como sus frecuencias dentro de la población. En la presente tesis se han considerado la Entropía de Shannon (H_s), la frecuencia de mutación máxima ($Fmáx$) y la diversidad de nucleótidos (π).

A continuación se proporciona una definición de cada uno de estos índices:

- **Número de sitios polimórficos (M), número de mutaciones (nM), número de haplotipos (Hpl) y densidad de haplotipos ($dHpl$)**

M : número de sitios nucleotídicos diferentes en los que tiene lugar algún evento de mutación en un conjunto de lecturas N

nM : número total de mutaciones diferentes que hay en un conjunto de lecturas N

Las posibles diferencias entre M y nM se deben a que algunas posiciones experimentan sustituciones diferentes.

Hpl : número de haplotipos o secuencias diferentes en un conjunto de lecturas N

$dHpl$: densidad de haplotipos calculada como Hpl/N

- **Frecuencia de mutación mínima ($Fmín$)**

Da cuenta de la proporción de nucleótidos mutados en un conjunto de lecturas de tamaño N , contando solo una vez cada mutación, independientemente de que esta se repita en diferentes lecturas.

$$Fmín = \frac{nM}{N \cdot L}$$

siendo L la longitud de los genomas evaluados.

- **Entropía de Shannon (H_s)**

Mide la diversidad basada en la frecuencia de haplotipos distintos. Estrictamente hablando, es la medida de la incertidumbre en la asignación de un haplotipo determinado a una muestra aleatoria de la cuasiespecie viral (Zhao y Illingworth, 2019).

$$H_s = - \sum p_i \log(p_i)$$

donde p_i es la frecuencia del haplotipo i dentro de una muestra de N lecturas.

El valor de H_s normalmente se normaliza respecto al número de lecturas, N , o respecto al de haplotipos, H_{pl} , resultando los índices H_sN y H_sH respectivamente.

$$H_sN = H_s / \log(N)$$

$$H_sH = H_s / \log(H_{pl})$$

En ambos casos, el valor máximo, 1, se alcanza cuando todas las secuencias de la muestra son diferentes y el mínimo, 0, si todas son iguales.

- **Frecuencia de mutación máxima ($F_{máx}$)**

Indica la proporción de nucleótidos mutados en un conjunto de lecturas de tamaño N , contando todas las veces que se repite cada mutación.

$$F_{máx} = \sum \frac{p_i}{L}$$

siendo p_i la frecuencia de cada mutación en un conjunto de lecturas N , siendo L la longitud de cada una de ellas .

Las frecuencias de mutación máxima y mínima pueden ser calculadas respecto a la secuencia del virus ancestral ($F_{máx_{WT}}$ y $F_{mín_{WT}}$) o respecto a la secuencia consenso de cada población ($F_{máx_{SC}}$ y $F_{mín_{SC}}$). Salvo que se indique lo contrario, se mostrarán las determinadas respecto a la secuencia consenso.

- **Diversidad de nucleótidos (π)**

La diversidad de nucleótidos o diversidad nucleotídica π se define como la medida del promedio de número de nucleótidos diferentes entre N genomas dados. Para esto, se evalúan las diferencias por pares alélicos en cada posición del genoma secuenciado.

En la posición genómica l , donde se observan n_i copias del alelo i y n_j copias del alelo j , la proporción de las diferencias por pares entre los alelos, en un conjunto de N secuencias, D_l , se puede calcular como:

$$D_l = \frac{N(N-1) - \sum_i n_i(n_i-1)}{N(N-1)}$$

Donde $i=A, C, G, T/U$. De este modo, para un conjunto de N secuencias, π se puede calcular para un genoma o región genómica de longitud L como:

$$\pi = \sum_{i=1}^L \frac{D_l}{L}$$

Para el cálculo de todos estos índices de diversidad se implementó un algoritmo en lenguaje de programación C.

5. RESULTADOS

5.1 CAPÍTULO 1

Caracterización de poblaciones del bacteriófago Q β que difieren en el grado de diversidad genética

Con objeto de disponer de dos poblaciones del bacteriófago Q β que pudieran ser utilizadas como ancestros en los experimentos diseñados para analizar el efecto de la diversidad genética preexistente en la adaptación, se partió de un clon biológico obtenido a partir de la expresión de un cDNA del genoma viral clonado en el plásmido pBR322 (plásmido pBRT7Q β , ver Materiales y Métodos). Este clon, cuya secuenciación confirmó que no contenía ninguna mutación respecto a la secuencia del virus presente en pBRT7Q β , fue considerado un análogo del virus salvaje y se denominó Q β_{WT} (WT = wild type).

El virus Q β_{WT} se propagó a 37 °C durante 2 o 25 pases seriados, en las condiciones descritas en el apartado 4.2.2. de Materiales y Métodos, para dar lugar a las poblaciones ancestrales Q β_{P2} y Q β_{P25} respectivamente (Fig. 13). Durante el proceso de multiplicación del virus a lo largo de la serie de pases es esperable que se produzcan mutaciones, sobre las que actuará la selección natural, aumentando la representación de las que aporten más ventajas en las condiciones ensayadas (temperatura, medios de cultivo y procedimientos utilizados en nuestro laboratorio). Existen, por tanto, dos fuerzas contrapuestas actuando: la generación de diversidad debida a los errores de la replicación viral y la eliminación de parte de esa diversidad como consecuencia de la acción de la selección natural.

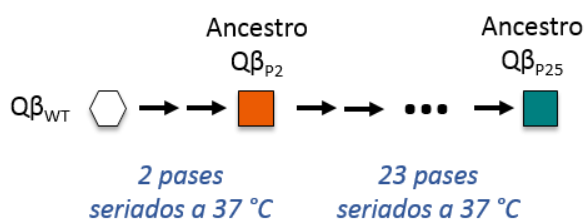


Figura 13: Historia evolutiva de las poblaciones ancestrales Q β_{P2} y Q β_{P25} .

La determinación de la secuencia consenso de ambas poblaciones mostró que Q β_{P2} no contenía ninguna mutación respecto a la secuencia del virus Q β_{WT} , mientras que Q β_{P25} había adquirido dos polimorfismos: A2187C (S281R en la proteína A1) y C3065U (una mutación sinónima en el gen de la replicasa).

A continuación se analizó si la propagación del virus a 37 °C había aumentado su capacidad replicativa a 43 °C. Para ello se realizó un ensayo de *fitness* a 43 °C, en el cual

comparamos las tasas de crecimiento de ambas poblaciones ancestrales (ver Materiales y Métodos). Los resultados obtenidos mostraron un valor de 4.7 ± 0.1 para $Q\beta_{P2}$ y de 4.8 ± 0.7 para $Q\beta_{P25}$, los cuales no fueron significativamente diferentes ($p > 0.1$, test t -Student).

5.1.1 Análisis mediante NGS de las poblaciones $Q\beta_{P2}$ y $Q\beta_{P25}$

Para analizar si, como consecuencia del diferente número de pases experimentado, las poblaciones $Q\beta_{P2}$ y $Q\beta_{P25}$ diferían en su estructura intrapoblacional y en el grado de diversidad genética que contenían, se extrajo RNA de ambas y se sometió a un protocolo de RT-PCR (PCR1) seguido de secuenciación masiva (proyecto Sec1; ver apartado 4.2.6. de Materiales y Métodos) que permitiera analizar un número elevado de lecturas correspondientes a tres regiones del genoma viral (Fig. 14). Estas regiones se escogieron de modo que incluyeran algunas de las posiciones nucleotídicas que habían experimentado mutaciones en experimentos previos de adaptación de $Q\beta$ a temperatura elevada. Así, en la región genómica del amplicón 1 (nucleótidos del 1060 al 1331) se localizan las sustituciones A1088G, G1312A, A1295C y A1295G; en el amplicón 2 (nucleótidos del 2145 al 2473) se localizan A2187C, G2223A y C2228U y en el amplicón 3 (nucleótidos del 3329 al 3603) U3402C y C3545U. El número de lecturas totales obtenido para cada amplicón, así como las conservadas tras el procesado bioinformático (ver el apartado 4.2.6. de Materiales y Métodos) se muestra en la Tabla 6. Es remarcable que en todos los casos el número de lecturas disponibles para los análisis posteriores fue siempre superior a 200000.

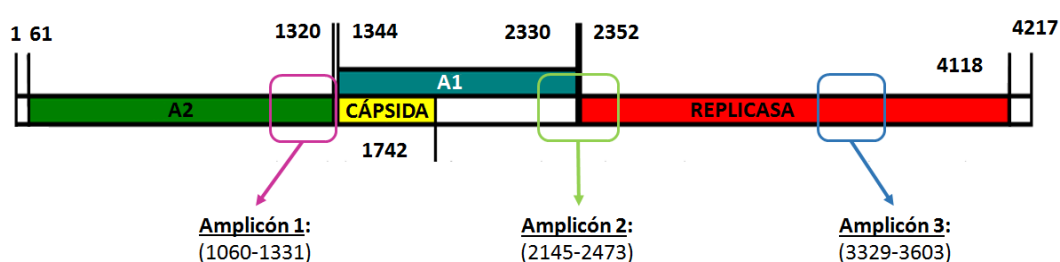


Figura 14: Mapa mostrando las regiones genómicas del bacteriófago $Q\beta$ analizada por secuenciación masiva.

Distribución de las secuencias más abundantes en cada amplicón

La distribución de las abundancias de las 100 secuencias (también llamadas haplotipos a partir de ahora) más frecuentes en el conjunto de lecturas obtenido para cada uno de los tres amplicones de las poblaciones $Q\beta_{P2}$ y $Q\beta_{P25}$ se muestra en la Fig. 15. En dicha figura también se indica la posición de tres umbrales, escogidos arbitrariamente, que incluyen los haplotipos representados en al menos el 0.05 %, el 0.1 % o el 0.5 % del total

de las lecturas (Tabla 7). En ambas poblaciones se observa que existen unos pocos haplotipos muy representados y una cola muy larga de haplotipos que están presentes en abundancias muy bajas y que se prolongaría bastante más allá de los 100 mostrados en la figura (Fig. 15; Tabla 7). Como consecuencia, la disminución del umbral de representación incrementa en mucha mayor proporción el número de haplotipos que el de lecturas.

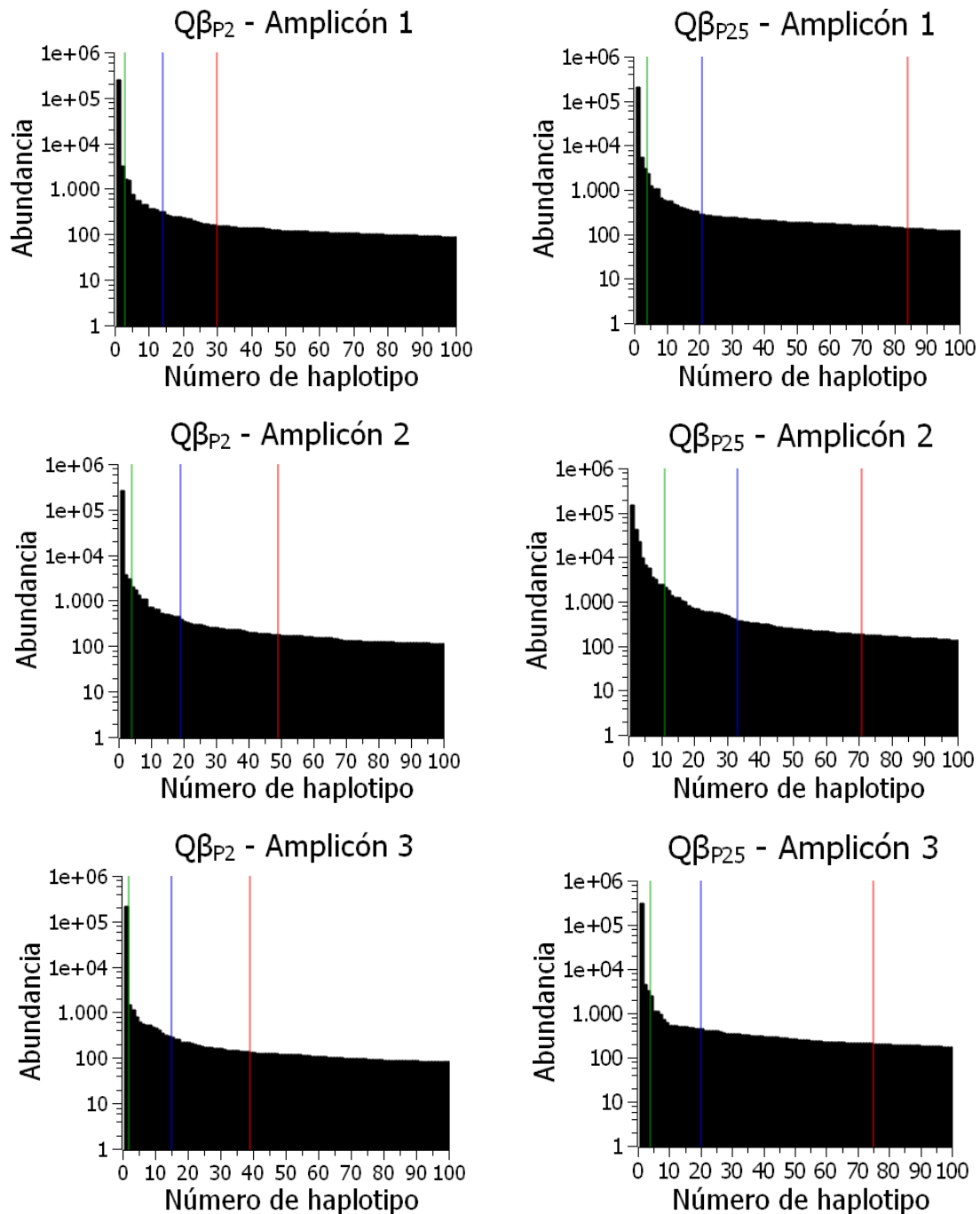


Figura 15: Distribución de las 100 secuencias más frecuentes en las poblaciones $Q\beta_{P2}$ y $Q\beta_{P25}$ para cada uno de los amplicones en estudio. En el eje x se indica el número de orden de los diferentes haplotipos en relación a su abundancia en la población. Las líneas verticales indican los umbrales de abundancia escogidos (verde: 0.5 %, azul: 0.1 % y rojo: 0.05 %).

		Umbral 0 %		Umbral 0.05 %		Umbral 0.1 %		Umbral 0.5 %	
		N_0	Hpl	N	Hpl	N	Hpl	N	Hpl
Amp1	Q β _{P2}	309124	13688	266858	30	263529	14	257548	3
	Q β _{P25}	274466	14710	235849	84	223901	21	214541	4
Amp2	Q β _{P2}	358986	26127	294855	49	287631	19	276471	4
	Q β _{P25}	359406	31345	279452	71	269959	33	253148	11
Amp3	Q β _{P2}	262456	15078	223334	39	219111	15	212404	2
	Q β _{P25}	408717	20728	351353	75	336017	20	326161	4

Tabla 7: Número de lecturas (N) y de haplotipos (Hpl) obtenidos en la secuenciación masiva de las poblaciones Q β _{P2} y Q β _{P25}. (Amp1: amplicón 1, Amp2: amplicón 2 y Amp3: amplicón 3). El umbral 0 % incluye todas las secuencias obtenidas tras el procesado bioinformático (N_0), mientras que el resto de umbrales incluyen solo las que están representadas por encima del valor indicado.

Determinación del umbral de representación de las secuencias utilizado en los análisis posteriores

Para analizar qué fracción de las lecturas obtenidas era más útil para nuestros análisis, se realizó un experimento control en el que el RNA extraído de la población Q β _{P25} fue amplificado en una nueva reacción de RT-PCR (PCR2) y secuenciado en un nuevo proyecto de secuenciación (Sec2), en el que se analizaron las mismas regiones genómicas mostradas en la Fig. 14. El resultado de la primera RT-PCR (PCR1), correspondiente a los amplicones mostrados en la Tabla 7 y la Fig. 15, también se volvió a secuenciar en el proyecto Sec2. Como resultado, se obtuvieron las combinaciones PCR1/Sec1, PCR1/Sec2 y PCR2/Sec2 (Fig. 16).

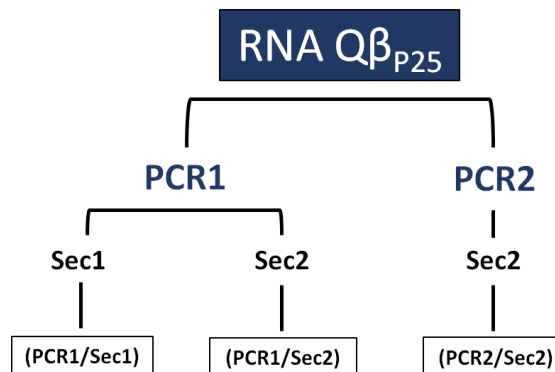


Figura 16: Esquema de las diferentes secuenciaciones de la población Q β _{P25}. Los resultados de la primera RT-PCR (PCR1), secuenciada en el proyecto Sec1, (PCR1/Sec1), son los mostrados en la Tabla 7 y la Fig. 15.

Puesto que en todos los casos se partió del mismo RNA viral, la determinación de las coincidencias entre las secuencias de haplotipos presentes en muestras diferentes permitirá estimar hasta qué punto los cambios genéticos encontrados corresponden a mutaciones reales producidas durante la replicación del virus o a errores producidos durante la RT-PCR y/o la secuenciación. El número exacto de lecturas consideradas para cada umbral y amplicón, así como el número de haplotipos y mutaciones totales identificados en las muestras PCR1/Sec1, PCR1/Sec2 y PCR2/Sec2 se muestran en la Tabla 8.

		Umbral 0.05 %			Umbral 0.1 %			Umbral 0.5 %		
		<i>N</i>	<i>Hpl</i>	<i>nM</i>	<i>N</i>	<i>Hpl</i>	<i>nM</i>	<i>N</i>	<i>Hpl</i>	<i>nM</i>
Amplicón 1	(PCR1/Sec1)	235849	84	81	223901	21	20	214541	4	3
	(PCR1/Sec2)	244723	83	81	233489	26	25	221899	4	3
	(PCR2/Sec2)	236742	83	81	224966	24	23	213660	3	2
Amplicón 2	(PCR1/Sec1)	279452	71	56	269959	33	25	253148	11	7
	(PCR1/Sec2)	206416	81	62	199197	43	32	182214	11	7
	(PCR2/Sec2)	203939	77	62	197492	43	31	182438	12	8
Amplicón 3	(PCR1/Sec1)	351353	75	74	336017	20	19	326161	4	3
	(PCR1/Sec2)	240322	81	80	230252	27	26	220849	4	3
	(PCR2/Sec2)	236158	80	79	226501	29	28	216817	4	3

Tabla 8: Número de lecturas (*N*), haplotipos (*Hpl*) y mutaciones (*nM*) obtenidos para cada amplicón en cada uno de los procesos de secuenciación de Q β _{P25}. Se muestran los resultados obtenidos para tres umbrales de representación de los haplotipos.

De acuerdo con lo mostrado en las distribuciones de abundancias (Fig. 15), al descender el umbral se está dando cabida en el análisis a haplotipos que están cada vez menos representados. En todos los casos, el número de haplotipos incluidos para el umbral mayor considerado (0.5 %) es muy bajo, lo cual es en cierto modo esperable dado que la pequeña longitud de las tres regiones analizadas hace que la mayoría de las lecturas no contengan mutaciones, salvo en la región 2 en la que aparece la mutación A2187C como polimorfismo claramente identificable en la secuencia consenso.

Se observa que, cuando se consideran los umbrales del 0.1 y el 0.5 %, el amplicón 2 contiene mayor número de haplotipos que el 1 o el 3. Sin embargo, cuando se considera el umbral más bajo (0.05 %), la inclusión de un número mayor de secuencias poco representadas disminuye esas diferencias.

Para todos los umbrales, los amplicones 1 y 3 muestran un número de haplotipos solo ligeramente superior al número de mutaciones, lo que indica que casi todos se diferencian en una única mutación. Sin embargo, en el amplicón 2 la diferencia es mayor,

probablemente debido a que muchos de ellos tienen ya la mutación A2187C, la cual puede aparecer combinada con otras.

A continuación se analizaron las coincidencias entre los haplotipos presentes en diferentes muestras. Para ello, para cada amplicón, se comprobó si los haplotipos de la muestra que contenía un número menor de los mismos estaban o no representados en aquella con la que se comparaba. De ese modo se obtuvo un porcentaje de coincidencia de haplotipos por pares de muestras (Tabla 9).

Muestras comparadas	Umbral 0.05 %			Umbral 0.1 %			Umbral 0.5 %		
	Amp1	Amp2	Amp3	Amp1	Amp2	Amp3	Amp1	Amp2	Amp3
PCR1/Sec1 vs PCR1/Sec2	86.75%	92.96%	90.67%	95.24%	100.00%	80.00%	100.00%	100.00%	100.00%
PCR1/Sec1 vs PCR2/Sec2	84.34%	85.91%	82.67%	85.71%	93.94%	70.00%	100.00%	100.00%	100.00%
PCR1/Sec2 vs PCR2/Sec2	92.77%	92.21%	90.00%	95.83%	93.02%	74.07%	100.00%	100.00%	100.00%

Tabla 9: Porcentajes de coincidencia entre haplotipos. (Amp1: amplicón 1, Amp2: amplicón 2 y Amp3: amplicón 3).

Para el umbral mayor considerado (0.5 %), aplicando el criterio indicado, en todos los casos se observa un porcentaje de coincidencia del 100 %. Es de destacar que todos los haplotipos que se quedan fuera de la comparación, por no aparecer por encima del umbral del 0.5 % en una de las dos muestras, pudieron ser detectados al decrecer ligeramente el umbral de representación. A pesar de los buenos resultados en términos de coincidencia de haplotipos, considerar solo los que están representados por encima del 0.5 % nos permite incluir solo una fracción muy pequeña del total, lo que hace que no sea muy útil para el tipo de análisis que queremos realizar.

Los porcentajes de coincidencia entre haplotipos también son bastante altos cuando se consideran los umbrales del 0.05 % y 0.1 %, siendo los valores menores en el caso del amplicón 3 y el umbral del 0.1 %. El hecho de que al disminuir el umbral al 0.05 % aumente el porcentaje de coincidencia en esta región, sugiere que haplotipos que no coincidían por no estar representados por encima del 0.1 % en ambas poblaciones, sí lo están al aumentar el número de lecturas.

Como era de esperar, los valores de identidad más bajos se obtienen cuando muestras que han sido amplificadas en reacciones distintas son secuenciadas en proyectos diferentes. Cuando la misma muestra se secuencia en proyectos distintos o diferentes productos de PCR se secuencian simultáneamente se obtienen valores de coincidencia similares. La comparación más útil para distinguir las mutaciones producidas durante la replicación del virus de los errores de la secuenciación es la que confronta los haplotipos de la misma muestra secuenciada en proyectos diferentes. En este caso, puesto que con

el umbral del 0.05 %, el porcentaje de coincidencia entre haplotipos fue mayor del 85 % en las tres regiones genómicas, se escogió este umbral para todos los análisis de los resultados de NGS incluidos en esta tesis doctoral. A partir de ahora, los resultados de secuenciación masiva mostrados para la población $Q\beta_{P25}$ se referirán siempre a la muestra PCR1/Sec1.

Distribuciones de haplotipos

Para comprobar si el diferente número de generaciones experimentado por las poblaciones $Q\beta_{P2}$ y $Q\beta_{P25}$ conduce a cambios en su estructura, en primer lugar se analizaron las distribuciones de frecuencias de los haplotipos representados por encima del 0.05 % (Fig. 17).

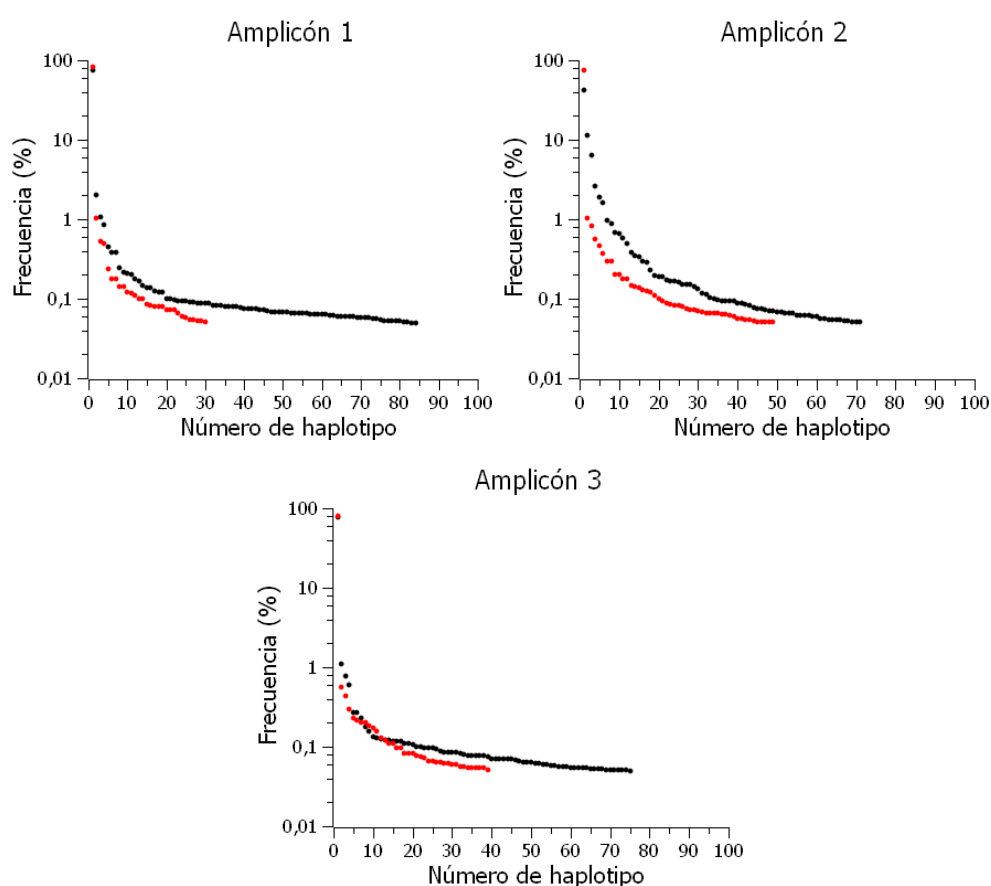


Figura 17: Frecuencias de los haplotipos representados en más del 0.05 % de las lecturas obtenidas. $Q\beta_{P2}$ (●), $Q\beta_{P25}$ (●). El número de cada haplotipo representa el número de orden en relación a su frecuencia.

Se observa que, en todos los amplicones, la frecuencia de los distintos haplotipos decae más rápidamente en $Q\beta_{P2}$ que en $Q\beta_{P25}$. Además, en esta última población, el rango de haplotipos es más amplio, como ya se mostraba en la Tabla 7. En el amplicón 2 de la población $Q\beta_{P25}$, el haplotipo más representado es el que contiene la mutación A2187C.

En el resto de casos, el haplotipo más representado es el que no contiene mutaciones respecto al virus Q β _{WT}.

A continuación se representó la distribución de frecuencias de los distintos haplotipos en función del número de mutaciones que contienen respecto a una secuencia de referencia (distancia de Hamming), que puede ser la del virus Q β _{WT} o la secuencia consenso de la población (Figs. 18 y 19).

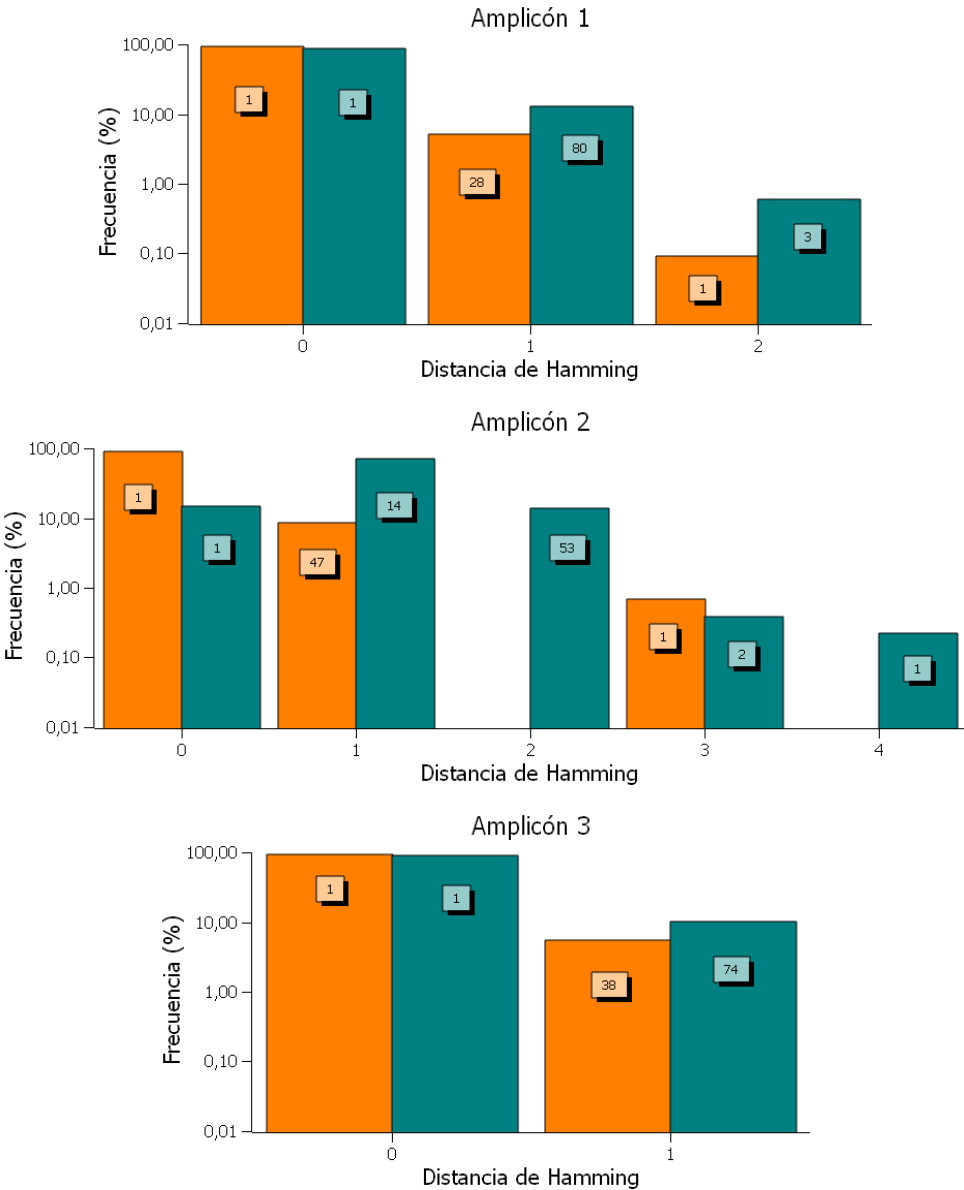


Figura 18: Frecuencia de haplotipos en función de su distancia de Hamming respecto al virus Q β _{WT} para las poblaciones Q β _{P2} (●) y Q β _{P25} (●). Las etiquetas sobre las barras muestran el número de haplotipos diferentes que se encuentran a la distancia de Hamming indicada. El número de haplotipos totales en cada amplicón, con representación en más del 0.05 % de las lecturas, se indica en la Tabla 7.

En todos los casos se observa que la representación del haplotipo que no contiene mutaciones (distancia de Hamming 0) es menor en Q β _{P25} que en Q β _{P2}, al mismo tiempo

que la frecuencia y el número de haplotipos a distancias de Hamming mayores o iguales a 1 es mayor. La diferencia entre las poblaciones es mucho más llamativa en el amplicón 2, en el que, en el caso de $Q\beta_{P25}$, la frecuencia de haplotipos a distancia de Hamming 1 se hace mayor que la del haplotipo sin mutaciones, lo cual es debido a la alta frecuencia del haplotipo conteniendo la mutación A2187C.

La representación de la distribución de haplotipos en función de la distancia de Hamming respecto a la secuencia consenso de cada población (Fig. 19) solo difiere de la mostrada en la figura anterior en el caso del amplicón 2 de $Q\beta_{P25}$, ya que en el resto de muestras la secuencia consenso coincidía con la del virus $Q\beta_{WT}$. En este caso, la forma de la distribución de $Q\beta_{P25}$ se asemeja más a la de $Q\beta_{P2}$ que cuando las diferencias se calculaban respecto a $Q\beta_{WT}$ (Fig. 18). Nuevamente el haplotipo a distancia de Hamming 0 está más representado en $Q\beta_{P2}$, mientras que los situados a distancias mayores o iguales que 1 son más frecuentes en $Q\beta_{P25}$.

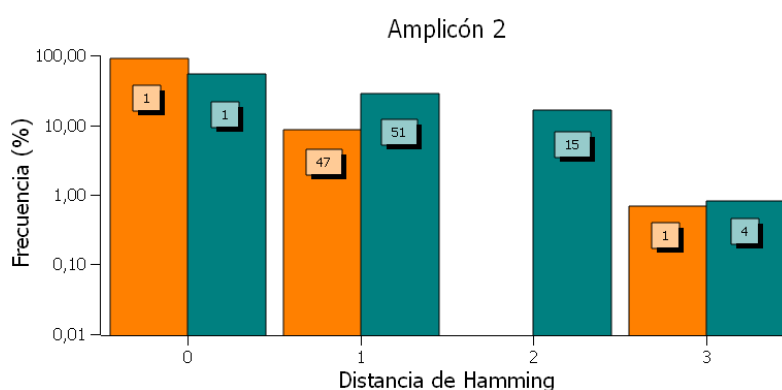


Figura 19: Frecuencia de haplotipos en función de su distancia de Hamming respecto a la secuencia consenso para las poblaciones $Q\beta_{P2}$ (●) y $Q\beta_{P25}$ (●). Las etiquetas sobre las barras muestran el número de haplotipos diferentes que se encuentran a la distancia de Hamming indicada. El número de haplotipos totales en cada amplicón, con representación en más del 0.05 % de las lecturas, se indica en la Tabla 7.

Identificación de las mutaciones y los haplotipos más representados en las poblaciones $Q\beta_{P2}$ y $Q\beta_{P25}$

Las dos poblaciones ancestrales $Q\beta_{P2}$ y $Q\beta_{P25}$ han experimentado cierto número de pases a 37 °C, durante los cuales se han generado mutaciones que han sido expuestas a la acción de la selección natural. Para analizar si algunas de esas mutaciones había alcanzado cierta representación que pudiera ser relevante en la posterior adaptación a nuevas presiones selectivas, se procedió a identificar las que estaban representadas en una proporción mayor del 0.5 % en el conjunto de lecturas analizadas en cada amplicón (Tabla 10).

Q β _{P2}			Q β _{P25}		
	Mutación	Frecuencia (%)		Mutación	Frecuencia (%)
Amplicón 1	U1295G	0.67	Amplicón 1	A1088G	0.52
	G1312A	0.70		U1295C	1.35
	U1328G	1.19		U1295G	0.52
Amplicón 2	A2187C	1.00		G1312A	2.85
	C2189U	1.27	Amplicón 2	U1328G	0.99
	G2211A	0.68		A2187C	68.80
	A2212U	0.68		C2189U	1.13
	U2213C	0.68		C2201U	6.21
	C2391A	0.57		G2217A	2.23
Amplicón 3	C3545U	0.67		G2223A	2.93
	A3603G	0.50		U2225C	1.69
				C2228U	8.62
				G2253A	1.11
				U2255C	0.87
			Amplicón 3	U3402C	0.90
				C3545U	1.29
				A3603G	0.71

Tabla 10: Mutaciones representadas en proporción mayor del 0.5 % en las poblaciones Q β _{P2} y Q β _{P25}. Las mutaciones sombreadas son las que están presentes por encima del 0.5 % en ambas poblaciones.

Se observa que la población Q β _{P25} posee mayor número de mutaciones por encima del umbral del 0.5 % que la población Q β _{P2} (17 frente a 11), de las cuales 7 están presentes en ambas poblaciones. Entre ellas destaca A2187C, que a pase 25 llega a alcanzar una frecuencia del 69 %. En ambas poblaciones, la mayor cantidad de mutaciones se detecta en el amplicón 2, lo cual sugiere que en esa región genómica se seleccionan más mutaciones con efectos beneficiosos en las condiciones de propagación del virus (entre ellas la temperatura de 37 °C) que en las otras dos analizadas. Esta región es también la que más cambios experimenta entre los pases 2 y 25, incluyendo la pérdida de 4 de las mutaciones detectadas en Q β _{P2} y la aparición de 7 mutaciones adicionales (C2201U, G2217A, G2223A, U2225C, C2228U, G2253A y U2255C). Respecto a las mutaciones presentes en las otras dos regiones genómicas, algunas de las presentes en Q β _{P2} incrementan ligeramente su representación en Q β _{P25} (G1312A, C3545U y A3603G), mientras que otras alcanzan el valor umbral del 0.5 % solo en esta última población (A1088G, U1295C y U3402C). El resto (U1295G, U1328G y C2189U) experimentan leves descensos en sus frecuencias.

Una observación intrigante es que 8 de las 17 mutaciones presentes por encima del 0.5 % en las poblaciones ancestrales propagadas a 37 °C, habían sido previamente identificadas (fijadas o como polimorfismos) en experimentos de adaptación de Q β a

43 °C (Arribas y col., 2014; 2016; Kashiwagi y col., 2014; Lázaro y col., 2018). Estas mutaciones son: A1088, U1295C, G1312A, A2187C, C2201U, G2223A, C2228U, U3402C. También es reseñable que durante la evolución de Q β a 37 °C en nuestro laboratorio, frecuentemente se ha fijado la mutación A2187C y las mutaciones A1088G y U1295C a veces han sido detectadas como polimorfismos (Arribas y col., 2014; 2016).

A continuación se analizaron las secuencias de los cinco haplotipos más representados en cada uno de los amplicones analizados en ambas poblaciones (Tabla 11). Como ya se ha indicado, excepto en el amplicón 2 de Q β _{P25}, el haplotipo más frecuente en todos los casos no contiene ninguna mutación, lo cual está de acuerdo con la corta longitud de cada amplicón y la ausencia de una presión selectiva fuerte que favorezca la selección de mutaciones. De hecho, entre los haplotipos mostrados en la Tabla 11 solo hay dos que contengan más de una mutación. En general, la mayoría de las mutaciones presentes en los haplotipos más representados coinciden con aquellas cuya frecuencia es mayor del 0.5 %.

		Amplicón 1		Amplicón 2		Amplicón 3	
		Q β _{P2}	Q β _{P25}	Q β _{P2}	Q β _{P25}	Q β _{P2}	Q β _{P25}
Haplotipo 1	Mutaciones	Ninguna	Ninguna	Ninguna	A2187C	Ninguna	Ninguna
	Frecuencia	94.71	86.38	90.81	54.97	94.44	89.93
Haplotipo 2	Mutaciones	U1328G	G1312A	C2189U	-	C3545U	C3545U
	Frecuencia	1.19	2.34	1.27	14.79	0.67	1.29
Haplotipo 3	Mutaciones	G1312A	U1295C	A2187C	C2228U	A3603G	U3402C
	Frecuencia	0.61	1.36	1.00	8.10	0.50	0.90
Haplotipo 4	Mutaciones	U1295G	U1328G	G2211A, A2212U, U2213C	A2187C, C2201U	A3566G	A3603G
	Frecuencia	0.58	0.99	0.68	3.41	0.35	0.71
Haplotipo 5	Mutaciones	U1327G	A1088G	C2391A	C2201U	G3341U	C3517U
	Frecuencia	0.28	0.52	0.57	2.42	0.27	0.32

Tabla 11: Frecuencias (%) y mutaciones de los 5 haplotipos más representados en las poblaciones ancestrales Q β _{P2} y Q β _{P25}. Las mutaciones se determinaron respecto a la secuencia del virus Q β _{WT}.

Determinación de la diversidad genética contenida en las poblaciones Q β _{P2} y Q β _{P25} mediante la estimación de índices de diversidad

Existen numerosos parámetros que pueden ser utilizados para determinar la diversidad contenida en una población a partir de las secuencias de los componentes individuales de la misma. Entre ellos, algunos están basados en la incidencia de entidades concretas, sin tener en cuenta la frecuencia de cada una de ellas (Gregory y col., 2016). Entre esas entidades las más frecuentes son las mutaciones y los haplotipos. Tras la correspondiente normalización por el número de nucleótidos analizados (en el caso de las mutaciones) o por el número de lecturas (en el caso de los haplotipos) se obtienen

los índices $F_{mín}$ (frecuencia de mutación mínima) y $dHpl$ (densidad de haplotipos). A su vez, la $F_{mín}$ se puede determinar respecto a la secuencia del virus fundador de las líneas ($F_{mín_{WT}}$), lo cual nos puede dar cierta idea de cómo se ha diversificado la población desde su origen, o respecto a la secuencia consenso ($F_{mín_{SC}}$). Sin embargo, dado que en las regiones consideradas, la secuencia consenso y la secuencia del virus ancestral como máximo difieren en una única mutación (A2187C en el amplicón 2), en la práctica las dos determinaciones de $F_{mín}$ dan valores muy similares y solo se presentarán los obtenidos respecto a la secuencia consenso, que serán denotados por $F_{mín}$.

Otros índices de diversidad que sí tienen en cuenta las frecuencias de las entidades consideradas son, entre otros, la frecuencia de mutación máxima ($F_{máx}$), la diversidad nucleotídica (π) y las diferentes formas de la entropía de Shannon (H_s , H_sH , H_sN). Al igual que sucedía con la frecuencia de mutación mínima, la frecuencia de mutación máxima puede ser calculada respecto a la secuencia del virus ancestral ($F_{máx_{WT}}$) o respecto a la secuencia consenso de cada población ($F_{máx_{SC}}$). $F_{máx_{WT}}$ puede proporcionar una estimación de la intensidad de la selección natural, ya que cuánto más alto es su valor, más probable es que mutaciones concretas estén altamente representadas. $F_{máx_{SC}}$ aporta una información diferente que guarda más relación con la estructura interna de la población que $F_{máx_{WT}}$. Incluso si la secuencia consenso y la del virus salvaje presentan pocas diferencias, $F_{máx_{SC}}$ podría diferir sensiblemente de $F_{máx_{WT}}$. En esta tesis solo se presentarán los valores de $F_{máx_{SC}}$, por lo que de ahora en adelante estos se denotarán como $F_{máx}$.

La definición exacta de todos los índices de diversidad utilizados en esta tesis doctoral se encuentra en el apartado 4.2.7. de Materiales y Métodos.

Índices de diversidad basados en incidencia

Los valores de $dHpl$ y $F_{mín}$ para el conjunto de secuencias representadas en una proporción mayor del 0.05 % de las lecturas obtenidas para $Q\beta_{P2}$ y $Q\beta_{P25}$ se muestran en la Tabla 12 y en la Fig. 20.

		N	Hpl	dHpl($\cdot 10^{-4}$)	nM	$F_{mín}(\cdot 10^{-7})$
Amplicón 1	$Q\beta_{P2}$	266858	30	1.12	28	3.86
	$Q\beta_{P25}$	235849	84	3.56	81	12.63
Amplicón 2	$Q\beta_{P2}$	294855	49	1.66	50	5.15
	$Q\beta_{P25}$	279452	71	2.54	56	6.09
Amplicón 3	$Q\beta_{P2}$	223334	39	1.74	38	6.19
	$Q\beta_{P25}$	351353	75	2.14	74	7.66

Tabla 12: Valores de los índices de diversidad basados en incidencia en las poblaciones ancestrales $Q\beta_{P2}$ y $Q\beta_{P25}$. La definición exacta de los parámetros mostrados en la tabla se encuentra en el apartado 4.2.7. de Materiales y Métodos.

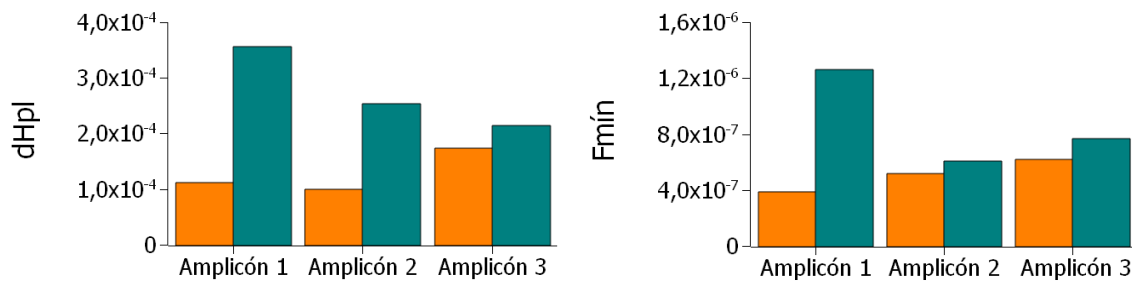


Figura 20: Densidad de haplotipos y frecuencia de mutación mínima en las poblaciones Qβ_{P2} (●) y Qβ_{P25} (●). La definición exacta de los parámetros mostrados en la figura se encuentra en el apartado 4.2.7. de Materiales y Métodos

En todos los casos, tanto *dHpl* como *Fmín* son mayores para la población Qβ_{P25} que para la población Qβ_{P2}, lo que indica que el mayor número de rondas de replicación experimentado por la primera favorece la presencia de un mayor número de mutaciones y de haplotipos diferentes, independientemente de cómo la selección natural moldee la representación de cada uno de ellos. El amplicón 1 es el que más diferencias presenta entre ambas poblaciones, un resultado claramente diferente del observado cuando se consideraba el número de mutaciones representadas en al menos el 0.5 % de las lecturas, cuyo número alcanzaba el valor mayor en el amplicón 2 (Tabla 10).

Índices de diversidad basados en abundancia

Los valores de *Fmáx*, π , *Hs*, *HsH* y *HsN*, para el conjunto de secuencias representadas en una proporción mayor del 0.05 % en Qβ_{P2} y Qβ_{P25}, se muestran en la Tabla 13 y la Fig. 21. Se observa que estos parámetros son siempre mayores para la población Qβ_{P25}, encontrándose las mayores diferencias en el amplicón 2.

		N	Hs	HsN($\cdot 10^{-2}$)	HsH	$\pi (\cdot 10^{-3})$	Fmáx($\cdot 10^{-4}$)
Amplicón 1	Qβ _{P2}	266858	0.36	2.88	0.11	0.39	1.98
	Qβ _{P25}	235849	0.91	7.37	0.21	1.04	5.23
Amplicón 2	Qβ _{P2}	294855	0.62	4.92	0.16	0.64	3.20
	Qβ _{P25}	279452	1.91	15.23	0.45	3.15	30.53
Amplicón 3	Qβ _{P2}	223334	0.40	3.26	0.11	0.40	2.02
	Qβ _{P25}	351353	0.72	5.61	0.17	0.73	3.66

Tabla 13: Valores de los índices de diversidad basados en abundancia en las poblaciones ancestrales Qβ_{P2} y Qβ_{P25}. La definición exacta de los parámetros mostrados en la tabla se encuentra en el apartado 4.2.7. de Materiales y Métodos

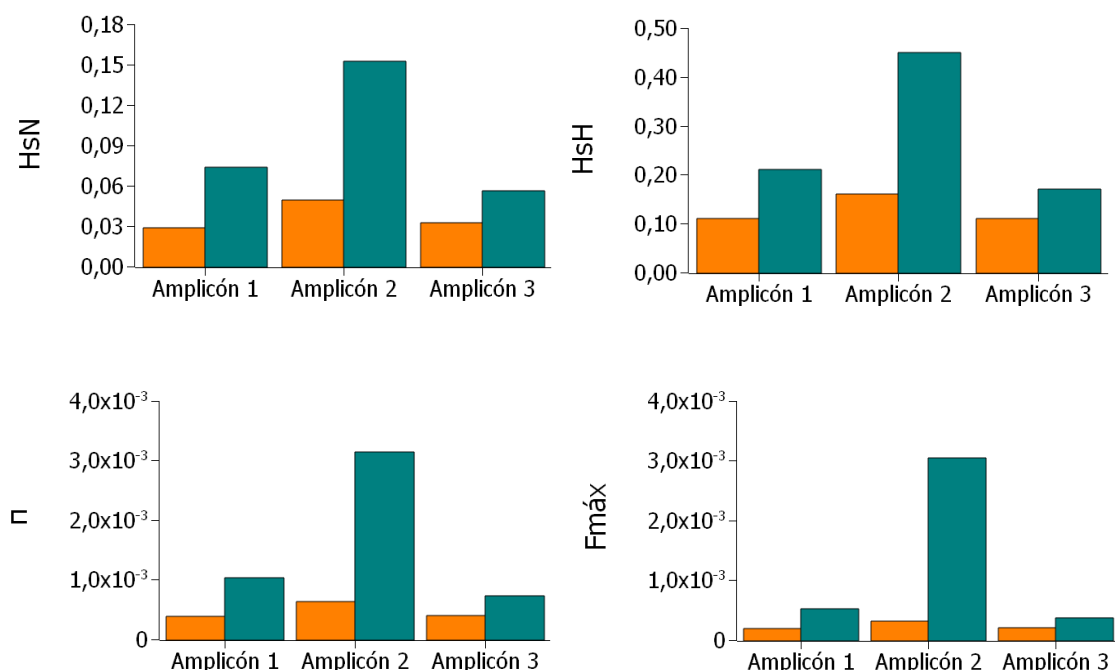


Figura 21: Valores de los índices de diversidad basados en abundancia. Poblaciones ancestrales QβP2 (●) y QβP25 (●). La definición exacta de los parámetros mostrados en la figura se encuentra en el apartado 4.2.7. de Materiales y Métodos

El aumento en el número de mutaciones representadas en más del 0.5 % de las lecturas que se observa en el amplicón 2 de QβP25 respecto a QβP2 (Tabla 10) podría ser una de las causas de los notorios aumentos en los valores de los índices de diversidad basados en abundancia observados en esta región genómica. En particular, la presencia de la mutación A2187C, que está presente en QβP25 en una proporción del 69 %, puede estar contribuyendo a alterar en un grado elevado la estructura de la población. El hecho de que *Fmín* y *dHpl* no sigan una tendencia similar, sino que muestren las mayores diferencias entre las dos poblaciones en el amplicón 1, puede indicar que muchas de las mutaciones que surgen en esta región se mantienen poco representadas a lo largo de la serie de pases, alterando en un grado menor la distribución de haplotipos.

5.1.2 Variación de los índices de diversidad en función del umbral de representación de los haplotipos

Para evaluar la robustez de nuestros resultados, se analizó si el umbral de representación de las secuencias escogido para el estudio podría afectar a los valores de los índices de diversidad de las poblaciones y a las diferencias observadas entre ellas. Para ello se compararon los valores obtenidos para los umbrales 0.05 %, 0.1 % y 0.5 % (Tablas 14 y 15).

		Umbral	N	Hpl	dHpl($\cdot 10^{-4}$)	nM	Fmín($\cdot 10^{-7}$)
Amplicón 1	Q β _{P2}	0.05	266858	30	1.12	28	3.86
		0.10	263529	14	0.53	13	1.81
		0.50	257548	3	0.12	2	0.28
	Q β _{P25}	0.05	235849	84	3.56	81	12.6
		0.10	223901	21	0.94	20	3.28
		0.50	214541	4	0.19	3	0.51
Amplicón 2	Q β _{P2}	0.05	294855	49	1.66	50	5.15
		0.10	287631	19	0.66	20	2.11
		0.50	276471	4	0.14	5	0.55
	Q β _{P25}	0.05	279452	71	2.54	56	6.09
		0.10	269959	33	1.22	25	2.81
		0.50	253148	11	0.43	7	0.84
Amplicón 3	Q β _{P2}	0.05	223334	39	1.75	38	6.19
		0.10	219111	15	0.68	14	2.32
		0.50	212404	2	0.09	1	0.17
	Q β _{P25}	0.05	351353	75	2.13	74	7.66
		0.10	336017	20	0.59	19	2.06
		0.50	326161	4	0.12	3	0.33

Tabla 14: Valores de los índices de diversidad basados en incidencia de las poblaciones Q β _{P2} y Q β _{P25} para los diferentes umbrales analizados. La definición exacta de los parámetros mostrados en la tabla se encuentra en el apartado 4.2.7. de Materiales y Métodos

Se observó que, tanto en Q β _{P2} como en Q β _{P25}, la densidad de haplotipos y la frecuencia de mutación mínima disminuyen al aumentar el umbral (Tabla 14). Las diferencias entre las dos poblaciones (Fig. 22) son mayores para el umbral menor considerado (0.05 %), algo que es particularmente evidente en el caso del amplicón 1.

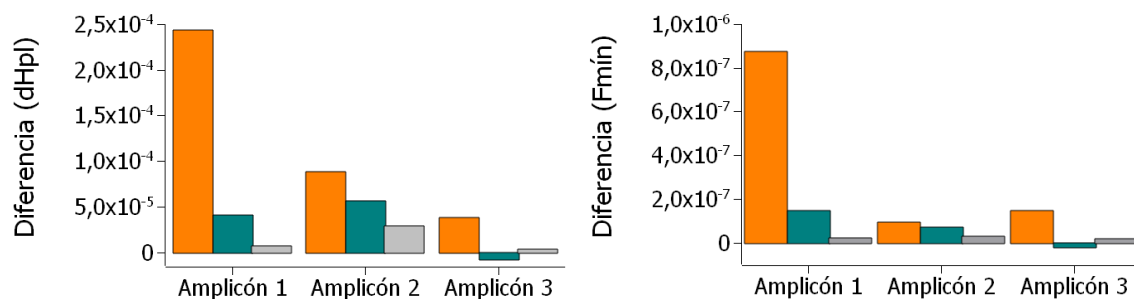


Figura 22: Diferencias entre poblaciones (Q β _{P25} - Q β _{P2}) para la densidad de haplotipos y la Frecuencia de mutación mínima. (●) umbral 0.05 %, (●) umbral 0.1 %, (●) umbral 0.5 %. La definición exacta de los parámetros mostrados en la figura se encuentra en el apartado 4.2.7. de Materiales y Métodos.

Los valores de los índices de diversidad basados en abundancia también tienden a disminuir en función del umbral de representación de las secuencias (Tabla 15), con la única excepción de H_sH en el caso del amplicón 2 de $Q\beta_{P25}$.

		Umbral	N	Hs	HsN ($\cdot 10^{-2}$)	HsH	π ($\cdot 10^{-3}$)	Fmáx($\cdot 10^{-4}$)
Amplicón 1	$Q\beta_{P2}$	0.05	266858	0.36	2.88	0.11	0.39	1.98
		0.10	263529	0.26	2.10	0.01	0.30	1.51
		0.50	257548	0.10	0.84	0.09	0.14	0.69
	$Q\beta_{P25}$	0.05	235849	0.91	7.36	0.21	1.04	5.23
		0.10	223901	0.53	4.29	0.17	0.69	3.48
		0.50	214541	0.25	2.05	0.18	0.36	1.85
Amplicón 2	$Q\beta_{P2}$	0.05	294855	0.62	4.9	0.16	0.64	3.20
		0.10	287631	0.43	3.43	0.15	0.50	2.52
		0.50	276471	0.17	1.38	0.12	0.28	1.39
	$Q\beta_{P25}$	0.05	279452	1.91	15.24	0.45	3.15	30.53
		0.10	269959	1.70	13.57	0.49	3.00	29.67
		0.50	253148	1.36	10.95	0.57	2.65	27.49
Amplicón 3	$Q\beta_{P2}$	0.05	223334	0.40	3.26	0.11	0.40	2.02
		0.10	219111	0.25	2.06	0.09	0.27	1.36
		0.50	212404	0.04	0.34	0.06	0.05	0.26
	$Q\beta_{P25}$	0.05	351353	0.72	5.61	0.17	0.73	3.66
		0.10	336017	0.38	2.98	0.13	0.43	2.17
		0.50	326161	0.17	1.36	0.12	0.22	1.14

Tabla 15: Valores de los índices de diversidad basados en abundancia de las poblaciones $Q\beta_{P2}$ y $Q\beta_{P25}$ para los diferentes umbrales analizados. La definición exacta de los parámetros mostrados en la tabla se encuentra en el apartado 4.2.7. de Materiales y Métodos.

Para todos estos parámetros, las diferencias entre las poblaciones $Q\beta_{P2}$ y $Q\beta_{P25}$ son mayores en el amplicón 2 y, con la única excepción de H_sH , se mantienen bastante constantes en esta región, independientemente del umbral (Fig. 23).

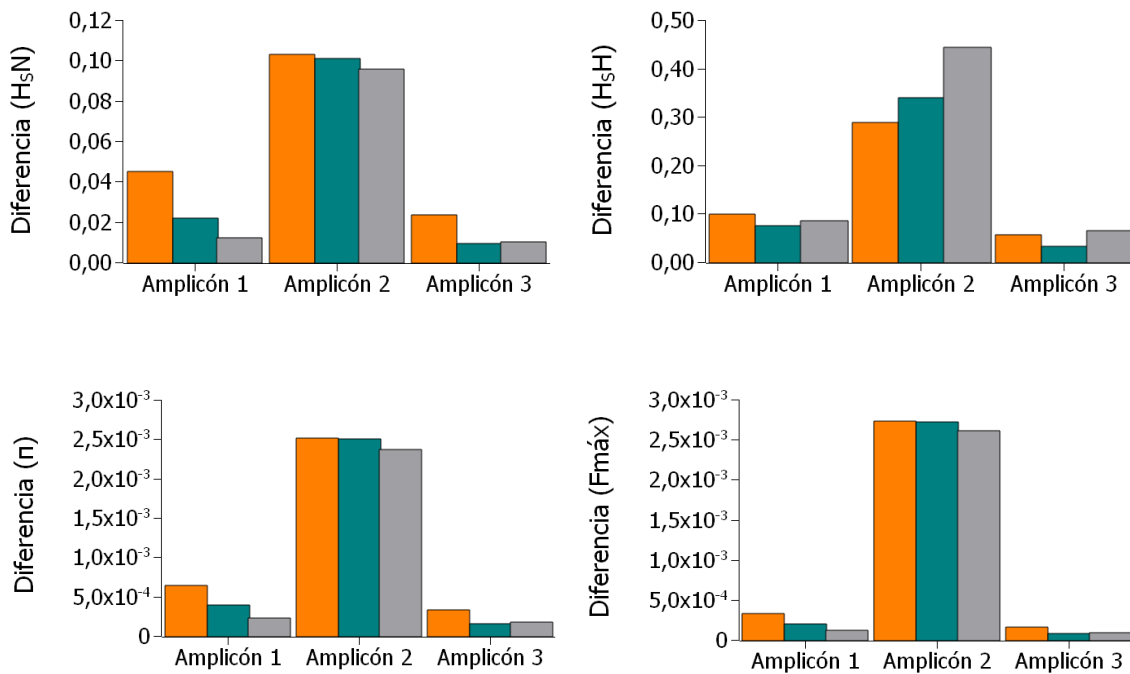


Figura 23: Diferencias entre poblaciones ($Q\beta_{P25} - Q\beta_{P2}$) para los valores de los índices de diversidad basados en abundancia. (●) umbral 0.05 %, (●) umbral 0.1 % y (●) umbral 0.5 %. La definición exacta de los parámetros mostrados en la figura se encuentra en el apartado 4.2.7. de Materiales y Métodos.

Considerando los amplicones que más diferencias presentan para los valores de los índices de diversidad de las poblaciones $Q\beta_{P2}$ y $Q\beta_{P25}$ (el 1 para los basados en incidencia y el 2 para los basados en abundancia) se observa que, en el caso del amplicón 1, las diferencias basadas en incidencia son muy dependientes del umbral, mientras que, en el caso del amplicón 2, las diferencias basadas en abundancia son más robustas. Esta observación puede deberse a que los índices de diversidad basados en abundancia están fuertemente determinados por los haplotipos más representados en la población, los cuales estarían incluidos en todos los umbrales (Gregori y col, 2016).

5.2 CAPÍTULO 2

Adaptación a temperatura elevada de las poblaciones del bacteriófago Q β que difieren en el grado de diversidad

Con el fin de analizar la influencia de la diversidad genética preexistente en la adaptación a una nueva presión selectiva, las poblaciones Q β_{P2} y Q β_{P25} descritas en el capítulo anterior se propagaron a 43 °C durante 60 pases seriados en triplicado (Fig. 24). De este modo, la población ancestral Q β_{P2} dio lugar a las líneas evolutivas C43 $_{P2.1}$, C43 $_{P2.2}$ y C43 $_{P2.3}$, mientras que, a partir de la población Q β_{P25} , se obtuvieron las líneas C43 $_{P25.1}$, C43 $_{P25.2}$ y C43 $_{P25.3}$. La notación C43 se escogió para indicar que ambas poblaciones evolucionaron de forma constante a 43 °C, mientras que los subíndices P2 y P25 hacen referencia a las poblaciones ancestrales Q β_{P2} y Q β_{P25} respectivamente.

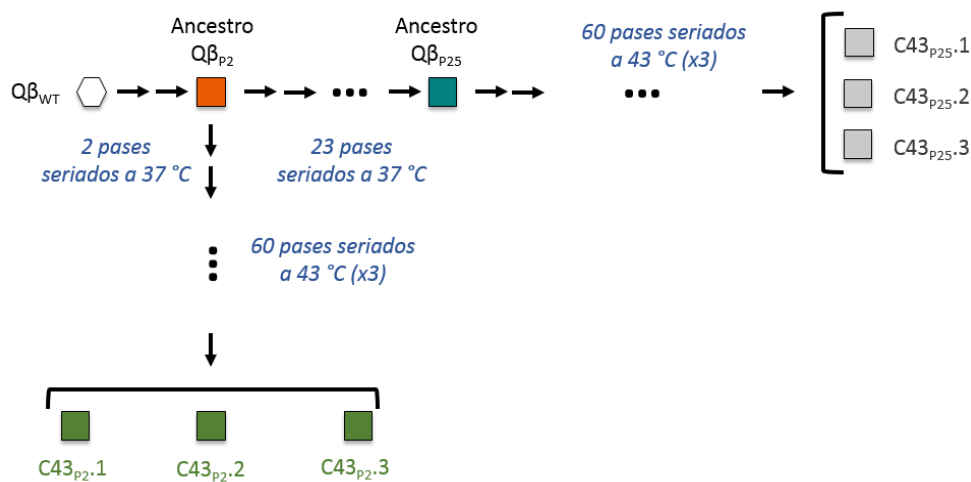


Figura 24: Esquema de obtención de las líneas evolutivas C43 $_{P2}$ y C43 $_{P25}$ a partir de sus ancestros correspondientes Q β_{P2} y Q β_{P25} .

5.2.1 Variación del *fitness* a lo largo de la adaptación

Para analizar la dinámica de la adaptación de Q β al aumento de temperatura, cada 10 pases se determinaron los valores de *fitness* absoluto a 43 °C. Estos valores se dividieron por el de la población ancestral correspondiente (4.7 ± 0.1 para Q β_{P2} y 4.8 ± 0.7 para Q β_{P25} ; ver Capítulo 1), para obtener así los valores de *fitness* relativo. Los resultados obtenidos se representan en la Fig. 25.

Como se puede observar, en el caso de las líneas evolutivas C43 $_{P25}$, la mayor parte del aumento de *fitness* tuvo lugar durante los primeros 10 pases, mientras que, en las líneas C43 $_{P2}$, el incremento fue más lento y de forma casi lineal hasta el pase 40.

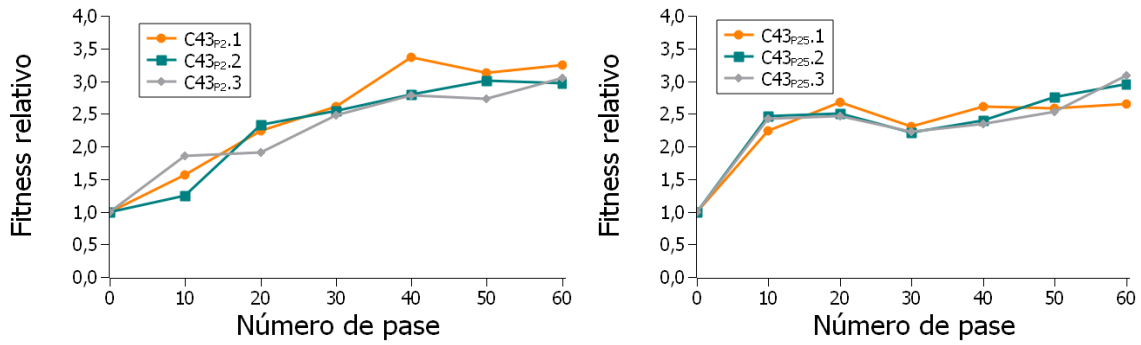


Figura 25: Dinámica de la variación del *fitness* a 43 °C en las líneas evolutivas C43P₂ y C43P₂₅.

Para comparar de forma cuantitativa las variaciones de *fitness* en las diferentes líneas evolutivas, se determinaron el número de pases necesarios para duplicar el valor de *fitness* de la población ancestral (es decir, para alcanzar un valor de *fitness* relativo de 2), los valores de *fitness* alcanzados en el último pase realizado (pase 60) y las diferencias de *fitness* obtenidas durante los primeros 10 pases a 43 °C. Todos estos valores de se muestran en la Tabla 16.

Línea evolutiva	Número de pase (fitness relativo = 2)	Fitness relativo a pase 60	Diferencia de fitness (primeros 10 pases a 43 °C)
C43P ₂ .1	16.3	3.3	0.6
C43P ₂ .2	16.9	3.0	0.3
C43P ₂ .3	21.5	3.1	0.9
C43P ₂ (Media ± SD)	18.2 ± 2.8	3.1 ± 0.1	0.6 ± 0.3
C43P ₂₅ .1	8.1	2.7	1.2
C43P ₂₅ .2	6.8	3.0	1.5
C43P ₂₅ .3	7.0	3.1	1.4
C43P ₂₅ (Media ± SD)	7.3 ± 0.7	2.9 ± 0.2	1.4 ± 0.1

Tabla 16: Parámetros cuantitativos que describen la dinámica de la variación del *fitness* a 43 °C en las líneas evolutivas C43P₂ y C43P₂₅. Media y SD hacen referencia a los valores medios y su correspondiente desviación estándar.

Aunque los valores de *fitness* alcanzados al final del experimento de adaptación, pase número 60, no presentaron diferencias significativas que pudieran ser atribuidas a la población ancestral ($p > 0.1$; test *t*-Student), sí se observaron profundas diferencias en el número de pases que fueron necesarios para doblar el *fitness* inicial ($p < 0.05$; test *t*-Student) y en las diferencias de *fitness* alcanzadas durante los diez primeros pases realizados a 43 °C ($p < 0.01$; test *t*-Student). La conclusión es que la población Q β _{P25} se adaptó más rápido a 43 °C que la población Q β _{P2}, aunque, después de un número prolongado de pases, ambas alcanzaron el mismo valor de *fitness*.

5.2.2 Dinámica de adquisición de mutaciones durante la adaptación a 43 °C

Cada diez pases se determinó la secuencia consenso de todas las líneas evolutivas y se identificaron las mutaciones presentes en las mismas, con el fin de analizar la dinámica de la adaptación a nivel genético. El genoma de referencia en todos los casos fue el del virus Q β _{WT}. La Tabla 17 muestra todas las mutaciones, fijadas y polimórficas, que se pudieron detectar, así como la estimación cualitativa de la intensidad de cada polimorfismo. Es importante destacar que cada posición mutada en una línea concreta fue examinada en el resto de líneas, para así identificar posibles polimorfismos que no hubieran sido detectados por los programas informáticos utilizados en el análisis.

Tabla 17.a: Líneas evolutivas C43_{P2}

	wt	mut<wt	mut=wt	mut>wt	mut
--	----	--------	--------	--------	-----

	C714U	A1088G	C1228U	U1295C/G	G1312A	G1371A	C1807U	C1810U	G1820A	U2006G	U2016C	A2187C	A2222C	G2223A
P10														
P20														
P30														
P40														
P50														
P60														

	C2452U	U2776C	U3311G	U3402C	U3447C	C3545U	C3903U	G3945A
P10								
P20								
P30								
P40								
P50								
P60								

Tabla 17.b: Líneas evolutivas C43_{P25}

	A1088G	C1228U	U1295C/G	G1312A	G1371A	U1544C	G1607U	C1649U	A1827G	A1956G	U2006G	U2016C	U2135C
P10													
P20													
P30													
P40													
P50													
P60													

	A2222C	C2384U	U2460C	U2641G	U2776C	G3002A	U3011C	U3311G	G3521U	G3896A	C3903U	G3945A
P10												
P20												
P30												
P40												
P50												
P60												

Tabla 17: Mutaciones y polimorfismos en las secuencias consenso de las líneas evolutivas C43_{P2} y C43_{P25}. La primera fila muestra los diferentes cambios nucleotídicos observados. En las filas siguientes aparece el número de pase analizado (P10 a P60). El estado de las mutaciones se muestra con un sombreado de diferente intensidad (ver parte superior de la tabla) en los cuadrados situados debajo de cada mutación (de izquierda a derecha, réplicas 1, 2 y 3 respectivamente). La posición 1295 experimentó cambios U/C o U/G que, con frecuencia, coexistían en la misma población. En la figura no se ha hecho distinción entre ambos tipos de cambio.

La representación del número total de mutaciones (fijadas y polimórficas) adquiridas por cada línea evolutiva en función del número de pases experimentados a 43 °C (Fig. 26), permitió analizar si la dinámica de adquisición de mutaciones era similar a la seguida por las variaciones de *fitness* (Fig. 25) y si había diferencias en función de la población ancestral.

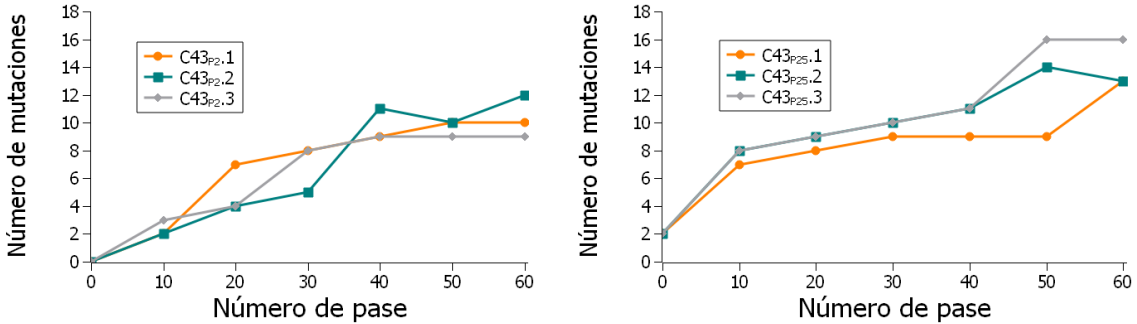


Figura 26: Número total de mutaciones (fijadas y polimórficas) en las secuencias consenso de las líneas C43_{P2} y C43_{P25} en función del número de pase.

Se observa que mientras que las líneas C43_{P2} adquirieron mutaciones de forma gradual hasta aproximadamente el pase número 40, las líneas C43_{P25} adquirieron mutaciones a mayor velocidad durante los diez primeros pases. Nuevamente, este resultado está en consonancia con las mayores ganancias de *fitness* experimentadas por estas líneas durante los primeros pases a 43 °C. La localización de las mutaciones en el genoma de cada población a pase número 60 se muestra en la Fig. 27. Se observa que la distribución no es uniforme, sino que hay regiones donde las mutaciones parecen acumularse de forma preferencial (parte final de la proteína A2 e inicial de la proteína de la cápsida, proteína A1 y algunas regiones de la replicasa), mientras que otras aparecen mucho más libres de mutaciones (parte central de la proteína A2).

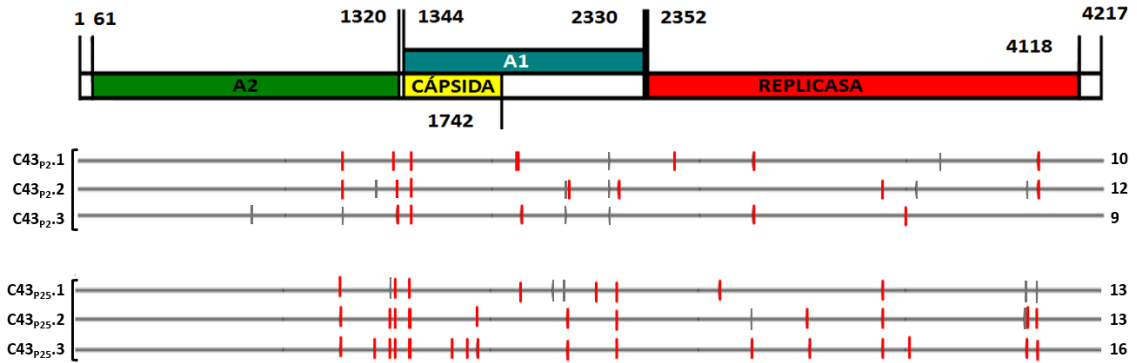


Figura 27: Mutaciones en las secuencias consenso de las líneas C43_{P2} y C43_{P25} en el pase número 60 de la adaptación a 43 °C. Las mutaciones fijadas en las poblaciones se representan con una línea roja y las mutaciones en estado polimórfico con una línea gris. A la derecha se indica el número de mutaciones totales en cada una de las líneas. La región secuenciada es la comprendida entre los nucleótidos 280 y 4150, según se indica en el apartado 4.2.5.

5.2.3 Comparación de las mutaciones adquiridas en los dos conjuntos de líneas evolutivas

El conjunto de todas las líneas evolutivas obtenidas a pase número 60 mostró un total de 33 mutaciones que afectaron 32 posiciones nucleotídicas (Tabla 18). Aunque la posición 1295 puede experimentar un cambio U/C o U/G, que implica una sustitución de fenilalanina por serina o cisteína respectivamente, por simplicidad y, dada la frecuente coexistencia de ambos cambios en la misma población, en este análisis no distinguiremos entre las dos situaciones.

Mutación ¹	Gen ²	Número de líneas con la mutación ³	
		C43 _{P2}	C43 _{P25}
C714U (S)	A2	0/1	
A1088G (D342G)	A2	2/1	3/0
C1228U (L389F)	A2	0/1	1/0
U1295(C/G) (F411S/C)	A2	1/0	2/1
G1312A (V417I)	A2	2/0	3/0
G1371A (G9S)	C	3/0	3/0
U1544C (S)	C		1/0
G1607U (Q87H)	C		1/0
C1649U (S)	C		2/0
C1807U (P154L)	A1	1/0	
U1810C (I155T)	A1	1/0	
G1820A (S)	A1	1/0	
A1827G (T161A)	A1		1/0
A1956G (K204E)	A1		0/1
U2006G (S220R)	A1	0/2	0/1
U2016C (F224L)	A1	1/0	2/0
U2135C (S)	A1		1/0
A2187C (S281R)	A1	0/3	
A2222C (S)	A1	1/0	3/0
C2452U (A33V)	R	1/0	
U2641G (L96R)	R		1/0
U2776C (V141A)	R	2/0	1/1
G3002A (S)	R		1/0
U3011C (S)	R		1/0
U3311G (I320M)	R	1/0	3/0
U3402C (S350P)	R	1/0	
U3447C (C365R)	R	0/1	
G3521U (E389D)	R		1/0
C3545U (S)	R	0/1	
G3896A (S)	R		0/1
C3903U (L517F)	R	0/1	2/1
G3945A (G531S)	R	2/0	2/1

Tabla 18: Mutaciones presentes en las líneas evolutivas C43_{P2} y C43_{P25} en el pase número 60 de la adaptación a 43 °C. ¹Todos los cambios se identificaron respecto a la secuencia del virus Qβ_{WT}. El cambio de aminoácido se muestra entre paréntesis. S indica que la mutación es sinónima. La posición 1295 puede mutar a C o G. ²Ubicación de la mutación en los diferentes genes del genoma de Qβ (C: cápsida, R: replicasa). ³El primer número de cada par indica el número de líneas en un conjunto dado en el que se fija la mutación. El segundo número se refiere al número de líneas en las que la mutación está presente como

polimorfismo. Las celdas vacías indican la ausencia de la mutación en las líneas de ese conjunto en particular.

Es de destacar que las mutaciones A2187C y C3065U, que estaban presentes como polimorfismos en la población ancestral Q β _{P25}, desaparecieron durante la evolución a 43 °C de las tres líneas obtenidas a partir de ella. No podemos saber si la pérdida de estas mutaciones fue debida a reversiones o a la selección negativa de los mutantes que las contienen. Es remarcable que A2187C apareció como polimorfismo en todas las líneas evolucionadas a partir de Q β _{P2}.

De las 32 mutaciones (considerando U1295C y U1295G como la misma) presentes a pase 60, 10 son sinónimas y 22 no sinónimas. Nueve de las mutaciones son exclusivas de las poblaciones C43_{P2}, 11 de las poblaciones C43_{P25} y 12 se encuentran en ambos conjuntos (Fig. 28). Es destacable que 11 de las 12 mutaciones comunes son no sinónimas, mientras que la proporción desciende a 11 de 20 en el caso de las mutaciones que solo están en uno de los dos conjuntos de líneas evolutivas. En general, las mutaciones exclusivas de C43_{P2} o C43_{P25} están presentes solo en una de las tres líneas evolutivas de cada conjunto (las únicas excepciones son las mutación A2187C que está presente como polimorfismo en las tres réplicas de C43_{P2} y la mutación C1649U que aparece fijada en dos de las réplicas de C43_{P25}). Sin embargo, las mutaciones comunes a ambos conjuntos presentan un comportamiento diferente en cada uno de ellos. Mientras que muchas de estas mutaciones (8 de 12) están presentes en las tres líneas evolutivas C43_{P25}, solo 2 aparecen en todas las líneas C43_{P2}.

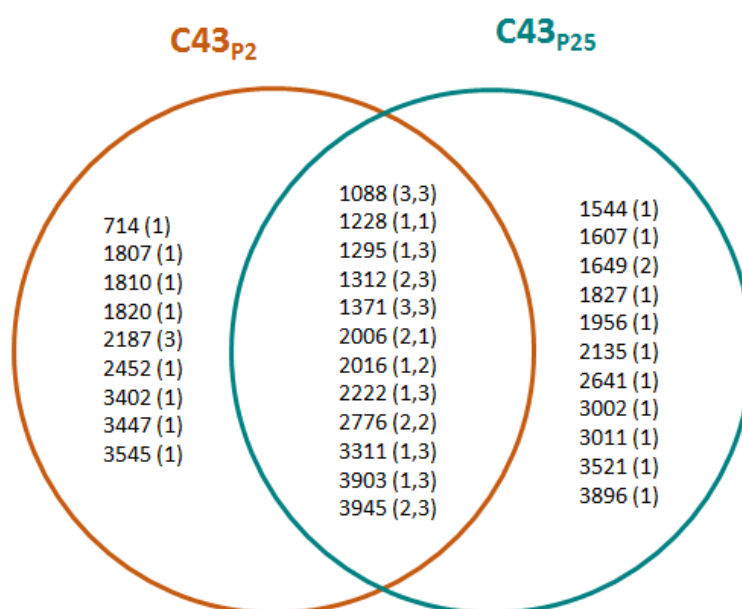


Figura 28: Diagrama de Venn mostrando las mutaciones coincidentes y exclusivas de las líneas evolutivas C43_{P2} y C43_{P25}. La circunferencia correspondiente a cada conjunto de líneas contiene las posiciones nucleotídicas que aparecen mutadas a pase 60 en al menos una línea. Entre paréntesis se indica el número de réplicas en las que la mutación está presente. En el caso de mutaciones coincidentes en los dos conjuntos, el primero de los números del paréntesis se refiere a las réplicas de C43_{P2} y el segundo a las de C43_{P25}.

Si se adopta el criterio de que las mutaciones más representadas en el conjunto de líneas analizado son las más relevantes para la adaptación a 43 °C, estas serían A1088G, U1295G/C, G1312A, G1371A, A2222C, U2776C, U3311G, C3903U y G3945A (están presentes al menos en 4 líneas). Es destacable que 3 de estas mutaciones (A1088G, U1295G/C y G1312A) ya estaban presentes en al menos una de las poblaciones ancestrales en una proporción mayor del 0.5 %, tal y como se mostró en el análisis de secuenciación masiva mostrado en el capítulo anterior (Tabla 10). Con la excepción de A2222C y U2776C, el resto de las mutaciones altamente representadas no están localizadas dentro de las regiones analizadas mediante secuenciación masiva y, por tanto, no se puede saber si estaban ya presentes en las poblaciones ancestrales.

Las mutaciones A2187C, U3402C y C3545U que aparecen en alguna de las líneas evolutivas C43_{p2} a pase 60, también fueron ya detectadas por encima del umbral del 0.5 % en la población ancestral Qβ_{p2}, aunque en el pase 60 no sean de las más representadas en el conjunto total de líneas.

Un punto importante cuando se estudian trayectorias adaptativas es la identificación de las mutaciones que se fijan primero. En el caso de las líneas analizadas, estas mutaciones se indican en la Tabla 19 observándose que, excepto C3903U y G3945A, todas las mutaciones representadas a pase número 60 en al menos cuatro líneas (ver arriba) están entre las primeras fijadas en al menos una línea, lo cual confirma su importancia en la adaptación a 43 °C. También parece que el primer evento de fijación de mutaciones ocurre más temprano en las líneas C43_{p25} que en las líneas C43_{p2}.

Línea evolutiva	Primeras mutaciones fijadas	Número de pase
C43_{p2}.1	A1088G, G1371A, C2452U, U2776C	30
C43_{p2}.2	A1088G, G1371A, A2222C, U3311G	20
C43_{p2}.3	G1312A, U3402C	10
C43_{p25}.1	A1088G, C1649U, U2016C, A2222C, U2460C, U3311G, G3945A	10
C43_{p25}.2	A1088G, A2222C, U3311G	10
C43_{p25}.3	A1088G, C1649U, A2222C, U3311G	10

Tabla 19: Primeras mutaciones fijadas durante la evolución a 43°C de las líneas C43_{p2} y C43_{p25}.
En la tercera columna de la tabla, se indica el número de pase en el que se detectó el primer evento de fijación de mutaciones.

Es de destacar que, con la única excepción de C43_{p2}.3, la mutación A1088G está entre las primeras fijadas en todas las líneas, haciéndolo de forma más temprana en las C43_{p25}, lo cual parece estar de acuerdo con su presencia por encima del 0.5 % en la población

ancestral $Q\beta_{P25}$. En la línea $C43_{P2.3}$, el primer evento de fijación de mutaciones ocurre a pases más tempranos que en las otras dos líneas con el mismo ancestro. En este caso las mutaciones que primero se seleccionaron fueron G1312A y U3402C, de las cuales G1312A también estaba presente en la población ancestral $Q\beta_{P2}$ con frecuencia mayor del 0.5 %.

5.2.4 Convergencia evolutiva entre líneas

Para determinar de forma cuantitativa el grado de similitud entre las vías evolutivas seguidas por las diferentes poblaciones, se analizó en primer lugar la diversidad nucleotídica entre las secuencias consenso de las líneas obtenidas a pase 60 procedentes de la misma población ancestral. Para ello se evaluó el número de diferencias nucleotídicas entre todos los pares posibles de secuencias consenso y se normalizó por el número de nucleótidos comparados. Los resultados mostraron que las líneas $C43_{P2}$ contenían mayor diversidad nucleotídica que las líneas $C43_{P25}$ (3.10×10^{-3} frente a 2.58×10^{-3}), lo cual está de acuerdo con la mayor coincidencia de mutaciones observada en el último caso (Tabla 17 y Tabla 18).

Además, se definieron varios índices de convergencia, que en todos los casos correspondieron a la relación entre el número de mutaciones comunes en las secuencias consenso de las líneas que comparten la misma población ancestral y el número de mutaciones totales en las secuencias consenso de esas mismas líneas (Tabla 18). Dependiendo de si solo se consideran las mutaciones fijadas o también las polimórficas y de si se cuentan como mutaciones comunes las que están presentes en todas las líneas o al menos en dos de ellas resultan cuatro índices de convergencia (Fig. 29).

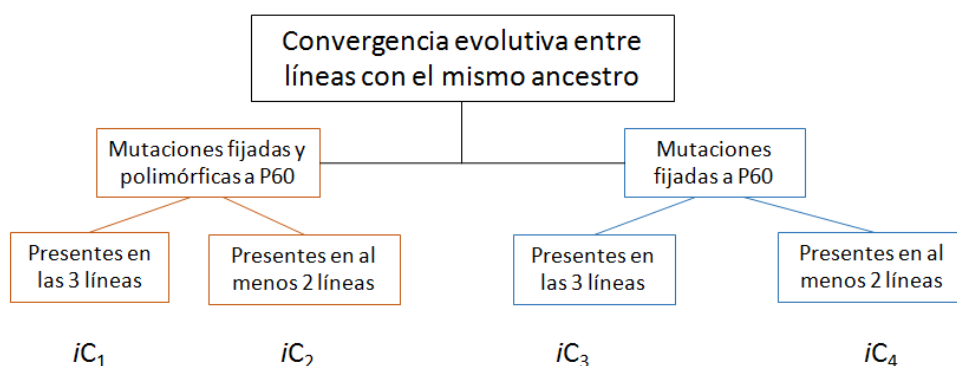


Figura 29: Mutaciones incluidas en el cálculo de los índices de convergencia de las líneas evolutivas $C43_{P2}$ y $C43_{P25}$. Si se tienen en cuenta las mutaciones fijadas y polimórficas de las poblaciones en estudio, se obtienen los índices de convergencia iC_1 (si las mutaciones comunes están presentes en las 3 líneas) e iC_2 (si las mutaciones comunes están en al menos 2 de ellas). Si solo se tienen en cuenta las mutaciones fijadas, se obtienen los índices iC_3 (si las mutaciones comunes están presentes en las 3 líneas) e iC_4 (si las mutaciones comunes están en al menos 2 de ellas).

Los valores obtenidos para cada uno de los índices de convergencia mostrados en la Fig. 29 se muestran en la Tabla 20.

	C43 _{P2}	C43 _{P25}
<i>iC₁</i>	0.14	0.35
<i>iC₂</i>	0.33	0.48
<i>iC₃</i>	0.07	0.25
<i>iC₄</i>	0.36	0.50

Tabla 20: Valores de los índices de convergencia para las poblaciones C43_{P2} y C43_{P25} en el pase 60 de la adaptación a 43 °C. Cada uno de estos valores ha sido calculado como se indica en la Fig. 29.

En todos los casos, las líneas C43_{P25} tienen valores más altos que las líneas C43_{P2}, lo que indica que después de 60 pases a 43 °C las líneas C43_{P25} presentan más similitud en las mutaciones que han adquirido a lo largo de la adaptación.

Los índices que solo consideran las mutaciones comunes que aparecen en 3 líneas (*iC₁* e *iC₃*) son más restrictivos, obteniéndose así valores más bajos que para los que consideran las mutaciones comunes que aparecen al menos en 2 de las 3 líneas (*iC₂* e *iC₄*).

La convergencia evolutiva entre los dos conjuntos de líneas C43_{P2} y C43_{P25} se determinó como el número de mutaciones presentes en al menos una línea de cada conjunto dividido entre el número total de mutaciones en los dos conjuntos. Si se consideran tanto las mutaciones fijadas como las polimórficas se obtiene el valor de 0.37.

5.3 CAPÍTULO 3

Estudio mediante NGS de la influencia de la diversidad genética preexistente en la adaptación a 43 °C

Para evaluar la influencia de la diversidad genética contenida en la población ancestral en la estructura de las nuevas poblaciones generadas durante la adaptación a 43 °C, se realizó un análisis mediante NGS de una línea representativa de cada conjunto C43_{P2} y C43_{P25}. Las líneas escogidas fueron la 3 en el caso de C43_{P2} y la 1 en el caso de C43_{P25}. Los procedimientos, regiones secuenciadas y umbral de representación de las secuencias utilizadas en el análisis fueron los mismos que en el caso de las poblaciones Q β _{P2} y Q β _{P25} (ver apartados 4.2.6. y 5.1.1.). En ambas líneas se secuenciaron el pase 1, para evaluar los cambios producidos inmediatamente después del aumento de temperatura a 43 °C, el pase 10 y el pase 60. El número de lecturas totales, las obtenidas tras el procesado bioinformático y las correspondientes a las secuencias representadas en más del 0.05 % del total de lecturas se muestran en la Tabla 6.

5.3.1 Distribuciones de haplotipos en las líneas C43_{P2.3} y C43_{P25.1}

Como una primera aproximación para analizar las diferencias entre las poblaciones adaptadas a 43 °C, se representaron las distribuciones de frecuencias de los haplotipos presentes en más del 0.05 % de las lecturas en el pase número 60. En la Fig. 30 se comparan las distribuciones de cada población evolucionada con las correspondientes a su población ancestral y en la Fig. 31 las dos poblaciones evolucionadas.

En las tres regiones genómicas analizadas, la población C43_{P2.3} se diferencia de su ancestro Q β _{P2} en la menor frecuencia del primer haplotipo y en la mayor representación de todos los restantes, lo cual hace que las curvas decaigan más rápidamente en Q β _{P2} que en C43_{P2.3}. También se observa un aumento claro en el número de haplotipos contenidos en las poblaciones. Todas estas observaciones son compatibles con la generación de una población más diversa a lo largo de la adaptación.

Los cambios experimentados por la población Q β _{P25} durante la adaptación a 43 °C son mucho menos notorios que los experimentados por la población Q β _{P2}. La frecuencia del haplotipo más representado en los amplicones 1 y 3 es menor en la población adaptada que en la ancestral, aunque las diferencias son menores que entre Q β _{P2} y C43_{P2.3}. En el amplicón 2, sin embargo, el haplotipo más representado está en mayor frecuencia tras la adaptación a 43 °C que en la población ancestral. En este caso también es destacable el cruce que se produce entre ambas distribuciones, mostrando que, a diferencia de lo que ocurre en el resto de casos, los haplotipos situados entre las posiciones 2 y 20 tienen frecuencias más bajas en la población adaptada que en la ancestral.

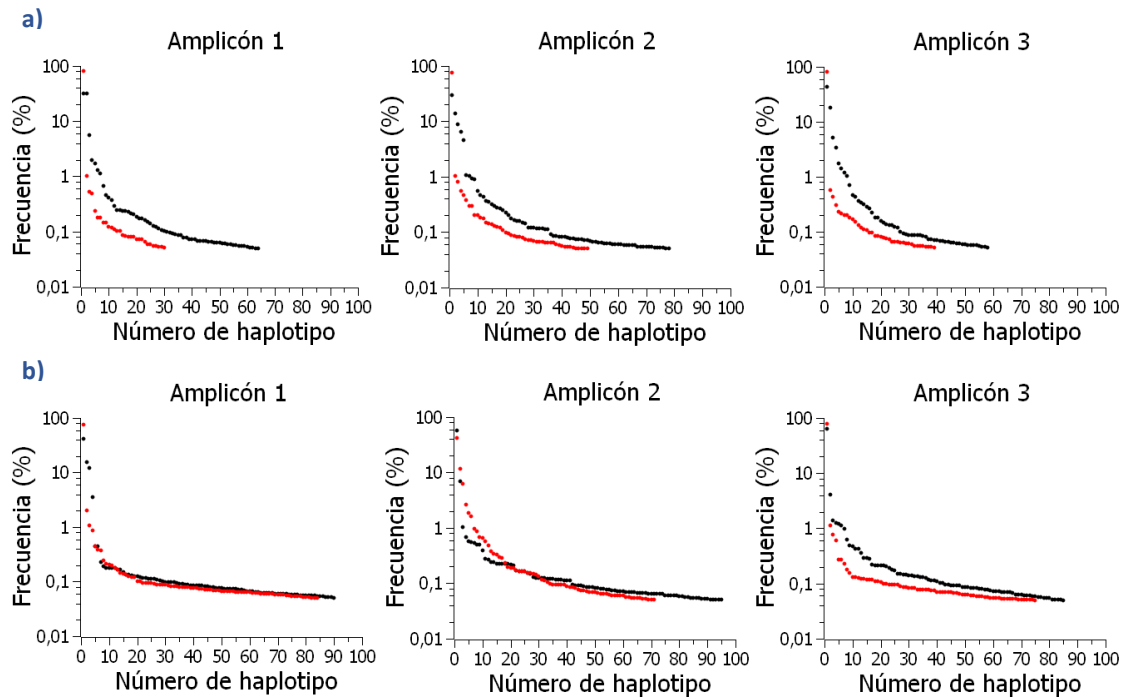


Figura 30: Frecuencias de los haplotipos representados en más del 0.05 % de las lecturas obtenidas a pase número 60 en cada uno de los amplicones analizados. a) $Q\beta_{P2}$ (●) y $C43_{P2.3}$ (●) b) $Q\beta_{P25.1}$ (●) y $C43_{P25.1}$ (●). El número de cada haplotipo representa el número de orden en relación a su frecuencia.

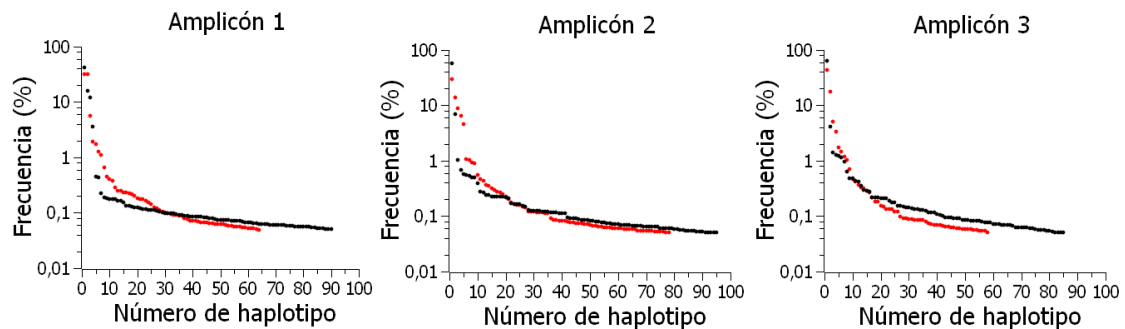


Figura 31: Frecuencias de los haplotipos representados en más del 0.05 % de las lecturas obtenidas a pase número 60 en cada uno de los amplicones analizados. $C43_{P2.3}$ (●) y $C43_{P25.1}$ (●). El número de cada haplotipo representa el número de orden en relación a su frecuencia.

La comparación de las dos poblaciones adaptadas a 43 °C (Fig. 31) muestra en todos los casos un punto de cruce entre ambas. En general, con la excepción del primer haplotipo, los haplotipos más frecuentes están en mayor proporción en la población $C43_{P2.3}$, mientras que los menos frecuentes lo están en la población $C43_{P25.1}$, la cual contiene además un número mayor.

La distribución de la frecuencia de haplotipos en función de su distancia de Hamming a la secuencia consenso se muestra en la Fig. 32. En todos los casos se observa que la frecuencia del haplotipo correspondiente a la secuencia consenso (distancia de Hamming 0) es mayor para la población $C43_{P25.1}$ que para la población $C43_{P2.3}$. Sin embargo, las frecuencias de los haplotipos a distancias de Hamming 1 o mayores son

claramente menores en el caso de la población C43_{P25.1}, todo lo cual parece indicar que esta población es más homogénea que la C43_{P2.3}. Este resultado contrasta con el hecho de que la población ancestral de C43_{P25.1} (Q β _{P25}) presentaba más diversidad que la de C43_{P2.3} (Q β _{P2}).

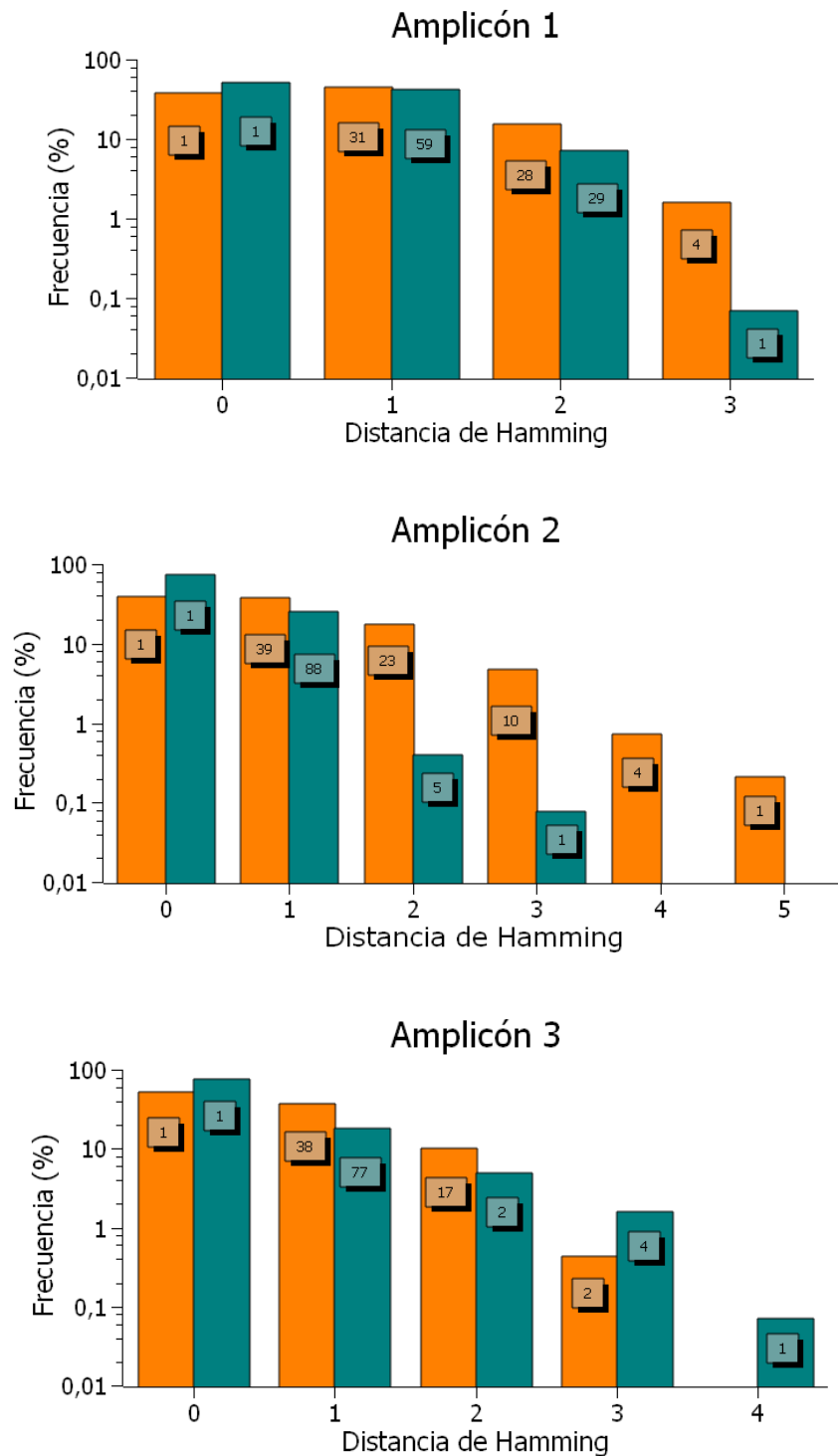


Figura 32: Frecuencia de haplotipos en función de su distancia de Hamming respecto a la secuencia consenso. Poblaciones C43_{P2.3} (●) y C43_{P25.1} (●) obtenidas a pase número 60. Las etiquetas sobre las barras muestran el número de haplotipos diferentes que se encuentran a la distancia de Hamming indicada. El número de haplotipos totales en cada amplicón, con representación en más del 0.05 % de las lecturas, se indica en la Tabla 6.

5.3.2 Evolución de la frecuencia de las mutaciones más representadas a lo largo de la adaptación a 43 °C

Para analizar cómo varía la frecuencia de las mutaciones presentes en las dos líneas evolutivas analizadas se establecieron dos umbrales de representación subjetivos para acotar las mutaciones de las que hacer un seguimiento. Como ya se indicó en el Capítulo 1, en las poblaciones ancestrales se consideraron como potencialmente relevantes todas aquellas que estaban representadas en más del 0.5 % del total de las lecturas analizadas. En las poblaciones evolucionadas a 43 °C se aumentó el umbral al 5 % para dar preferencia a las mutaciones cuya representación aumenta como consecuencia de la adaptación a 43 °C. De todas estas mutaciones se estimaron sus frecuencias en los pases 1, 10 y 60.

La variación a lo largo de la adaptación a 43 °C de las frecuencias de las mutaciones representadas en las poblaciones ancestrales por encima del 0.5 % se muestra en la Tabla 21. En la población ancestral Q β _{P2} existen 11 de estas mutaciones. Aunque ninguna de ellas está presente en una frecuencia elevada, en conjunto, pueden constituir parte del escenario sobre el que tendrá que comenzar a actuar la selección natural al producirse el incremento de temperatura a 43 °C. Después de un único pase a 43 °C, las frecuencias de las distintas mutaciones se mantienen en valores similares, mientras que a pase 10 ya es posible detectar un cambio muy notable que corresponde al incremento de la representación de la mutación G1312A hasta un valor del 82 %. Después de 60 pases a 43 °C, solo tres de las mutaciones presentes en más del 0.5 % de las lecturas de la población ancestral pasan a estar representadas por encima del umbral del 5 % en la población evolucionada (G1312A, A2187C, C3545U). No hay ninguna mutación que alcance el umbral del 5 % de forma transitoria en los pases que se han analizado.

En la población ancestral Q β _{P25}, se encuentran 17 mutaciones representadas por encima del 0.5 %. De ellas, solo cuatro (A1088G, U1295C, U1295G, G1312A) superan el umbral del 5 % a pase número 60. La mutación C3545U aumenta su representación respecto a la población ancestral, pero se mantiene por debajo del umbral del 5 %. El resto de mutaciones decrecen sus frecuencias (exceptuando algún ligero aumento en el pase número 1), lo que es especialmente evidente en el caso de las sustituciones A2187C, C2201U y C2228U que superaban el 5 % en la población ancestral, aunque solo A2187C había sido detectada como un polimorfismo mediante secuenciación Sanger.

		Frecuencia (%)			
	Mutación	Qβ _{P2}	C43 _{P2.3} (P1)	C43 _{P2.3} (P10)	C43 _{P2.3} (P60)
Amplicón 1	U1295G	0.67	0.56	0.58	3.17
	G1312A	0.70	0.56	81.51	91.45
	U1328G	1.19	0.74	0.95	0.58
Amplicón 2	A2187C	1.00	1.06	1.08	62.02
	C2189U	1.27	1.59	0.00	0.00
	G2211A	0.68	0.95	1.33	1.24
	A2212U	0.68	0.95	1.33	1.24
	U2213C	0.68	0.95	1.41	2.02
	C2391A	0.57	0.70	0.45	0.29
Amplicón 3	C3545U	0.67	0.61	0.46	8.82
	A3603G	0.50	1.19	0.97	0.59

		Frecuencia (%)			
	Mutación	Qβ _{P25}	C43 _{P25.1} (P1)	C43 _{P25.1} (P10)	C43 _{P25.1} (P60)
Amplicón 1	A1088G	0.52	0.93	99.07	99.65
	U1295C	1.35	1.36	0.29	15.56
	U1295G	0.52	0.57	0.39	58.45
	G1312A	2.85	3.63	15.04	99.63
	U1328G	0.99	0.87	0.66	0.50
Amplicón 2	A2187C	68.80	62.93	0.81	0.69
	C2189U	1.13	0.81	0.00	0.00
	C2201U	6.21	6.89	0.23	0.27
	G2217A	2.23	2.33	0.00	0.00
	G2223A	2.93	3.37	0.00	0.09
	U2225C	1.69	1.51	0.23	0.19
	C2228U	8.62	11.72	0.00	0.00
	G2253A	1.11	1.11	0.00	0.00
	U2255C	0.87	1.31	0.00	0.00
Amplicón 3	U3402C	0.90	1.50	0.66	0.16
	C3545U	1.29	0.89	0.96	3.14
	A3603G	0.71	1.06	0.95	0.74

Tabla 21: Trayectoria seguida por las mutaciones representadas en las poblaciones ancestrales en proporción mayor o igual al 0.5 % durante la evolución de C43_{P2.3} y C43_{P25.1}. Las mutaciones sombreadas en gris son las que a pase número 60 alcanzan una representación por encima del 5 %, mientras que las sombreadas en rosa corresponden a las que, estando representadas por encima del 5 % en la población ancestral, descienden por debajo de este valor a lo largo de la adaptación.

Es remarcable que ninguna de las mutaciones que estaban por encima del umbral del 0.5 % en la región 2 de $Q\beta_{P25}$ alcancen el umbral del 5 % tras la adaptación a 43 °C, lo cual contrasta con la presencia de algunas de ellas (A2187C, C2201U, G2223A y C2228U) en experimentos previos de adaptación a esta temperatura. Por el contrario, la mayoría de las mutaciones presentes en la región 1 de $Q\beta_{P25}$ (cuatro de cinco) alcanzan frecuencias elevadas después de 60 pases a 43 °C. La mayoría de los cambios descritos son ya observables a pase 10, mientras que, a pase 1, las frecuencias de las distintas mutaciones son muy similares a las observadas en la población ancestral, con las únicas excepciones de la mutación A2187C, que en solo un pase pasa del 69 % al 63 %, iniciando así su tendencia decreciente, y C2228U, que aumenta del 9 % al 12 %.

A continuación se procedió a identificar las mutaciones que no estaban presentes por encima del 0.5 % en la población ancestral y que alcanzan una proporción mayor del 5 % en alguno de los pases analizados durante el proceso de adaptación a 43 °C. Estas mutaciones se muestran en la Tabla 22, junto con sus frecuencias en cada pase.

	Mutación	Frecuencia (%)			
		$Q\beta_{P2}$	C43 _{P2.3} (P1)	C43 _{P2.3} (P10)	C43 _{P2.3} (P60)
Amplicón 1	A1088G	0.21	0.24	2.13	51.65
	U1295C	0.06	0.00	0.52	11.40
Amplicón 2	G2217A	0.25	0.06	0.08	8.51
	G2236A	0.07	0.07	5.51	0.00
	T2450C	0.00	0.00	0.00	14.16
	C2452U	0.00	0.08	0.34	13.11
Amplicón 3	U3402C	0.20	0.19	88.47	89.16
	G3494A	0.00	0.00	0.09	25.72

	Mutación	Frecuencia (%)			
		$Q\beta_{P25}$	C43 _{P25.1} (P1)	C43 _{P25.1} (P10)	C43 _{P25.1} (P60)
Amplicón 1	C1228U	0.00	0.00	0.19	5.08
Amplicón 2	C2210U	0.14	0.23	0.16	9.17
	A2222C	0.18	0.21	99.65	99.57
	U2460C	0.00	0.10	70.64	0.29
Amplicón 3	U3371C	0.00	0.07	0.10	6.70
	U3447C	0.00	0.06	2.63	8.06

Tabla 22: Mutaciones que no estaban presentes en las poblaciones ancestrales en más del 0.5 % de las lecturas y que alcanzan una representación mayor o igual al 5 % durante la evolución de las poblaciones C43_{P2.3} y C43_{P25.1}.

En la línea C43_{P2}.3 se detectan 8 nuevas mutaciones que alcanzan una representación mayor del 5 % a lo largo de la adaptación. Con la única excepción de G2236A, que alcanza el umbral del 5 % a pase 10 para decrecer después, todas presentan su máxima frecuencia a pase número 60. Es de destacar que 3 de esas 8 mutaciones (A1088G, U1295C y U3402C) sí que fueron detectadas en una proporción mayor del 0.5 % en la población ancestral Q β _{P25} y mostraron valores alrededor del 0.2 % en la población Q β _{P2}.

En la línea C43_{P25}.1 el número de nuevas mutaciones que alcanzan una representación mayor del 5 % es de 6, ninguna de las cuales coincide con las observadas en la línea C43_{P2}. Solo una de ellas (U2460C) supera el umbral del 5 % de forma transitoria, manteniéndose a pase 60 por debajo de esa frecuencia.

En conjunto, se observa que en las dos líneas evolutivas analizadas hay mayor coincidencia de mutaciones que superen el 5 % a pase número 60, entre las que ya estaban presentes por encima del 0.5 % en alguna de las poblaciones ancestrales (A1088G, U1295C, G1312A). Es interesante el comportamiento opuesto que muestra la mutación A2187C en ambas líneas evolutivas, aumentando claramente su representación en C43_{P2}.3 y disminuyéndola en C43_{P25}.1. Este hecho parece ser algo general, ya que la mutación A2187C aparece como polimorfismo de alta intensidad en la secuencia consenso de las tres líneas C43_{P2}, mientras que desaparece en todas las líneas C43_{P25} (Tabla 17).

Otra observación es que, además de las mutaciones fijadas o polimórficas que son claramente detectables en las secuencias consenso determinadas mediante secuenciación Sanger, a pase 60 también existen otras mutaciones, que están en proporciones apreciables (entre el 5 y el 26 %) y que no habíamos podido identificar en el análisis de las secuencias consenso. Algunas de estas mutaciones (C2452U, C3545U, U1295C) podrían ser relevantes para la adaptación a 43 °C, como se deduce del hecho de que en otros experimentos previos de adaptación a alta temperatura alcanzaron altos niveles o llegaron incluso a fijarse.

5.3.3 Variaciones en los haplotipos más representados a lo largo de la adaptación

Con objeto de analizar cómo se distribuyen las mutaciones mostradas en las Tablas 21 y 22 en los genomas virales, se determinaron las frecuencias de los cinco haplotipos más representados en los pases 1, 10 y 60 y los comparamos con los cinco más representados en la correspondiente población ancestral (Tablas 23 y 24).

		C43 _{P2}							
		Qβ _{P2}		C43 _{P2.3} (P1)		C43 _{P2.3} (P10)		C43 _{P2.3} (P60)	
Mutaciones		Posición	%	Posición	%	Posición	%	Posición	%
Amplícón 1	Ninguna	1	94.71	1	95.84	2	16.02	20	0.22
	U1328G	2	1.19	2	0.74	8	0.17	-	-
	G1312A	3	0.61	3	0.49	1	77.61	2	38.13
	U1295G	4	0.58	4	0.49	18	0.10	4	2.35
	U1327G	5	0.28	8	0.16	-	-	-	-
	A1088G	7	0.21	5	0.24	3	1.44	17	0.28
	A1088G,G1312A	-	-	-	-	5	0.69	1	38.58
	G1312A,U1328G	-	-	-	-	4	0.78	14	0.30
	A1088G,UT1295C	-	-	-	-	-	-	3	6.71
	A1088G,U1295C, G1312A	-	-	-	-	-	-	5	2.09
Amplícón 2	Ninguna	1	90.81	1	88.58	1	80.77	2	18.08
	C2189U	2	1.27	2	1.59	-	-	-	-
	A2187C	3	1.00	3	1.06	6	1.01	1	39.13
	G2211A,A2212U, U2213C	4	0.68	4	0.95	5	1.26	21	0.27
	C2391A	5	0.57	5	0.70	9	0.45	45	-
	G2236A	38	0.07	48	0.07	2	5.29	-	-
	C2228U	-	-	-	-	3	2.40	67	0.07
	G2223A	-	-	-	-	4	1.86	-	-
	U2450C	-	-	-	-	-	-	3	11.67
	A2187C,C2452U	-	-	-	-	-	-	4	8.48
	A2187C,G2217A	-	-	-	-	-	-	5	6.02
Amplícón 3	Ninguna	1	94.44	1	94.00	2	10.93	4	4.11
	C3545U	2	0.67	3	0.61	7	0.19	3	6.21
	A3603G	3	0.50	2	1.19	11	0.15	-	-
	A3566G	4	0.35	4	0.32	-	-	-	-
	G3341U	5	0.27	5	0.28	-	-	-	-
	U3402C	10	0.20	9	0.19	1	84.52	1	52.74
	U3402C,A3603G	-	-	-	-	3	0.82	13	0.41
	G3341U,U3402C	-	-	-	-	4	0.27	20	0.19
	U3402C,C3545U	-	-	-	-	5	0.27	5	2.10
	U3402C,G3494A	-	-	-	-	-	-	2	21.83

Tabla 23: Secuencias de los 5 haplotipos más representados a lo largo de la adaptación a 43 °C de la línea C43_{P2.3}. En cada columna aparece el número de orden de cada haplotipo y su frecuencia (%). Las mutaciones indicadas para cada haplotipo indican cambios respecto al virus Qβ_{WT}. Los cinco haplotipos más representados en cada pase aparecen sombreados. La posición de estos haplotipos también fue rastreada en el resto de los pases y, en el caso de haber sido localizados entre los contenidos en el umbral del 0.05 %, aparece indicada. Los casos en los que el haplotipo no ha sido encontrado en las poblaciones analizadas se denotan con -.

En la línea C43_{P2.3} (Tabla 23) se observa que las secuencias sin mutaciones en las tres regiones genómicas analizadas, que eran las más abundantes en la posición ancestral, descienden drásticamente su representación a lo largo de la adaptación. A pase 10, los haplotipos que más aumentan sus frecuencias son los que contienen G1312A en el amplicón 1 y U3402C en el amplicón 3. Las altas frecuencias alcanzadas por ambos (78 % y 85 %) sugieren que estas mutaciones están frecuentemente asociadas en el mismo genoma, lo cual también está de acuerdo con la secuencia consenso de la población a pase 10. A pase 60, el haplotipo más abundante cuando se considera el amplicón 1 contiene A1088G y G1312A. Puesto que estas mutaciones ya estaban presentes en la población ancestral, aunque en haplotipos separados, se plantea la cuestión de si su aparición simultánea en el mismo genoma se debe a recombinación o a su aparición recurrente. El haplotipo más representado en el amplicón 2, también a pase 60, es el que contiene la mutación A2187C y en el amplicón 3 el que contiene la mutación U3402C. En los tres amplicones se observa que el haplotipo más representado a pase 60 tiene menor frecuencia que el más representado a pase 10, algo que se traduce en un notable cambio en la distribución de haplotipos a lo largo de la adaptación (Fig. 33, parte superior).

Los resultados sugieren que la línea C43_{P2.3} el genoma mayoritario a pase 10 es el que contiene las mutaciones G1312A y U3402C, seguido del que contiene G2236A. El genoma mayoritario a pase 60 probablemente contiene G1312A, A1088G, A2187C, y U3402C, seguido de otro con las mutaciones G1312A, U3402C y G3494A. Obviamente, además pueden existir otras mutaciones que están localizadas en otras regiones genómicas no tenidas en cuenta en el análisis de NGS, tal y como se observa en las secuencias consenso (Tabla 17).

En la población ancestral Q β _{P25} (Tabla 24) el haplotipo mayoritario en los amplicones 1 y 3 era el que no contenía mutaciones. En cambio en la región 2, este haplotipo era el segundo más representado, por detrás del que contenía A2187C. A lo largo de la adaptación el haplotipo sin mutaciones pasa a ser casi indetectable en los amplicones 1 y 2, mientras que se mantiene como el más representado en el amplicón 3, donde a pase 60 todavía representa el 75.5 %. A pase 10, el haplotipo más representado en el amplicón 1 pasa a ser el que contiene A1088G, que ya estaba entre los cinco más representados en la población ancestral. Los tres siguientes en orden de frecuencia contienen esa misma mutación acompañada de otras. A pase número 60, los 5 haplotipos más representados poseen las mutaciones A1088G y G1312A, las cuales pueden estar acompañadas de otras adicionales, como U1295G en el más abundante. En el caso del amplicón 2, los cinco haplotipos más representados a pase 10 contienen la mutación A2222C, que en el caso del que ocupa la primera posición está acompañada de U2460C. A pase número 60, el haplotipo mayoritario contiene solo A2222C, mientras que el que contenía A2222C y U2460C pasa a ocupar la posición 15. Por último, en el amplicón 3 no parece favorecerse la selección de ninguna mutación en particular, puesto que, como se ha indicado, el haplotipo más representado, a gran distancia del resto, es el que no contiene ninguna mutación.

		C43 _{P25}							
		Q β _{P25}		C43 _{P25.1} (P1)		C43 _{P25.1} (P10)		C43 _{P25.1} (P60)	
Mutaciones		Posición	%	Posición	%	Posición	%	Posición	%
Amplícón 1	Ninguna	1	86.38	1	86.11	5	0.39	-	-
	G1312A	2	2.34	2	3.13	31	0.12	52	0.09
	U1295C	3	1.26	3	1.30	-	-	-	-
	U1328G	4	0.99	5	0.87	-	-	-	-
	A1088G	5	0.52	4	0.93	1	74.26	23	0.14
	A1088G,G1312A	-	-	-	-	2	17.61	2	19.37
	A1088G,C1209U	-	-	-	-	3	0.59	-	-
	A1088G,U1328G	-	-	-	-	4	0.51	-	-
	A1088G,U1295G, G1312A	-	-	-	-	56	0.07	1	51.01
	A1088G,U1295C, G1312A	-	-	-	-	-	-	3	14.85
	A1088G,C1228G, G1312A	-	-	-	-	-	-	4	4.37
	A1088G,G1122A, U1295G,G1312A	-	-	-	-	-	-	5	0.55
Amplícón 2	A2187C	1	54.97	1	57.80	-	-	-	-
	Ninguna	2	14.79	2	20.70	26	0.15	40	0.15
	C2228U	3	8.10	3	4.40	-	-	-	-
	A2187C,C2201U	4	3.41	4	0.60	-	-	-	-
	C2201U	5	2.42	5	0.49	-	-	-	-
	A2222C,U2460C	-	-	66	0.09	1	61.85	15	0.29
	A2222C	59	0.08	-	-	2	22.15	1	73.81
	U2174C,A2222C	-	-	-	-	3	4.71	22	0.22
	A2222C,U2383G, U2460C	-	-	-	-	4	0.64	-	-
	A2222C,U2381G, U2460C	-	-	-	-	5	0.53	-	-
	C2210U,A2222C	-	-	-	-	-	-	2	8.77
	A2222C,C2384U	-	-	-	-	53	0.09	3	1.32
	A2222C,U2426C	-	-	-	-	33	0.12	4	0.88
	A2222C,U2379G	-	-	-	-	28	0.14	5	0.72
Amplícón 3	Ninguna	1	89.93	1	90.01	1	83.64	1	75.47
	C3545U	2	1.29	4	0.89	3	0.96	3	1.64
	U3402C	3	0.90	2	1.50	6	0.66	33	0.16
	A3603G	4	0.71	3	1.06	5	0.95	8	0.74
	C3517U	5	0.32	6	0.22	7	0.47	9	0.57
	G3341U	6	0.31	5	0.31	14	0.24	38	0.14
	U3447C	-	-	50	0.06	2	2.63	4	1.48
	G3393A	-	-	-	-	4	0.95	-	-
	U3371C,U3447C	-	-	-	-	-	-	2	4.82
	C3464U	16	0.14	47	0.07	13	0.25	5	1.45

Tabla 24: Secuencias de los 5 haplotipos más representados a lo largo de la adaptación a 43 °C de la línea C43_{P25.1}. En cada columna aparece el número de orden de cada haplotipo y su frecuencia (%). Las mutaciones indicadas para cada haplotipo indican cambios respecto al virus Q β _{wr}. Los cinco haplotipos más representados en cada pase aparecen sombreados. La posición de estos haplotipos también fue rastreada en el resto de los pases y, en el caso de haber sido localizados entre los contenidos en el umbral del 0.05 %, aparece indicada. Los casos en los que el haplotipo no ha sido encontrado en las poblaciones analizadas se denotan con -.

Con los datos de frecuencias de los distintos haplotipos, se puede intuir que el genoma más representado a pase 10 probablemente contiene las mutaciones A1088G, A2222C y U2460C, seguido de otro con las mutaciones A1088G, G1312A y A2222C. Es probable que entre los pases 10 y 20 este segundo genoma, en el que también surgiría la mutación U1295G, pase a desplazar al primero, ya que a pase 60 el genoma más representado parece ser el que contiene las mutaciones A1088G, A1295G G1312A y A2222C. Esta sustitución del genoma mayoritario por otro se puede deducir también de las mutaciones presentes en las secuencias consenso, que entre los pases 10 y 20 muestran la desaparición de varias mutaciones (entre ellas la ya mencionada U2460C) que estaban altamente representadas a pase 10. Al igual que ocurría durante la adaptación de la población Q β _{P2}, en este caso también encontramos que mutaciones que estaban en haplotipos diferentes en la población ancestral aparecen en el mismo haplotipo a pases posteriores, algo que no sabemos si ocurre por recombinación o por la aparición recurrente de las mismas mutaciones. En los tres amplicones, el cambio experimentado en las frecuencias de los 5 haplotipos más representados, entre los pases 10 y 60, es menos notorio que en el caso de la población C43_{P2.3} (Fig. 33, parte inferior).

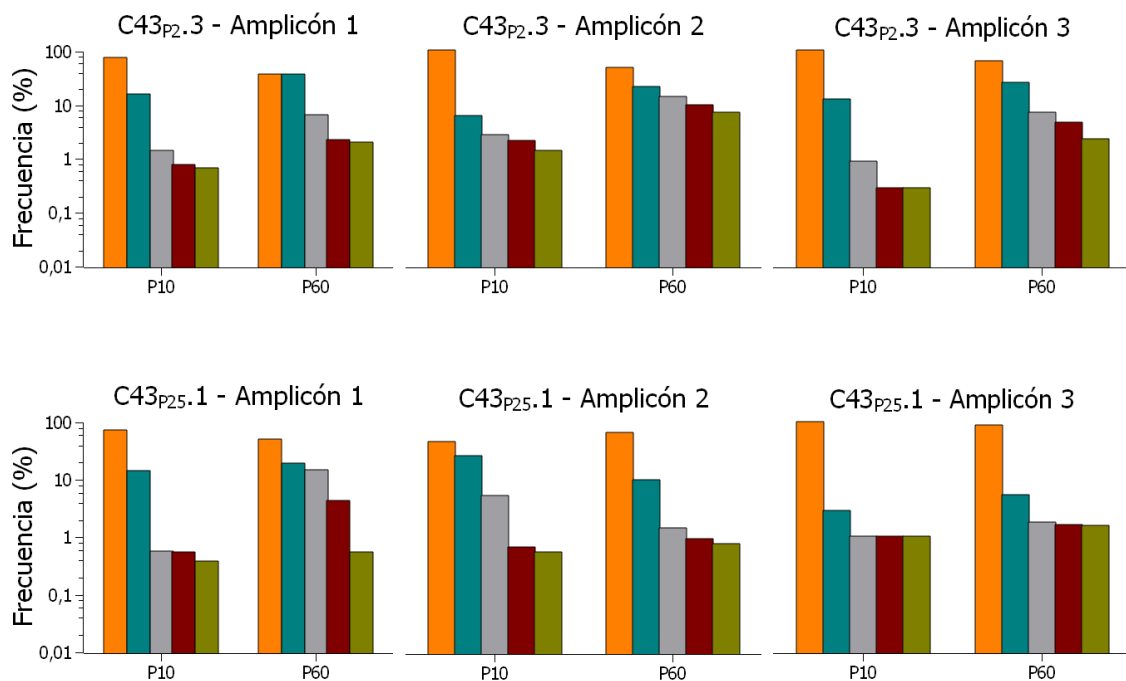


Figura 33: Frecuencia de los haplotipos más representados en las poblaciones C43_{P2.3} y C43_{P25.1} obtenidas a pase número 10 y 60. Representados el primero (●), segundo (●), tercero (●), cuarto (●) y quinto (●) haplotipos más abundantes en cada población y pases analizados.

5.3.4 Determinación de la diversidad genética en las poblaciones C43_{P2.3} y C43_{P25.1} mediante la estimación de índices de diversidad

Índices de diversidad basados en incidencia

La variación de los índices de diversidad basados en la incidencia de mutaciones, *F_{mín}*, y haplotipos, *dHp*, a lo largo de la adaptación a 43 °C de las líneas evolutivas C43_{P2.3} y C43_{P25.1} se muestra en la Tabla 25 y en la Fig. 28. En ambos casos también se incluyen los valores obtenidos para las poblaciones ancestrales, de modo que se pueda determinar su variación desde el origen.

		N	Hpl	dHp($\cdot 10^{-4}$)	nM	F _{mín} ($\cdot 10^{-7}$)
Amplicón 1	Q β _{P2}	266858	30	1.12	28	3.86
	C43 _{P2.3} (P1)	270956	28	1.03	26	3.53
	C43 _{P2.3} (P10)	239054	35	1.46	30	4.61
	C43 _{P2.3} (P60)	265606	64	2.41	32	4.43
Amplicón 2	Q β _{P2}	294855	49	1.66	50	5.15
	C43 _{P2.3} (P1)	274730	53	1.93	54	5.97
	C43 _{P2.3} (P10)	236250	52	2.20	50	6.43
	C43 _{P2.3} (P60)	252206	78	3.09	44	5.30
Amplicón 3	Q β _{P2}	223334	39	1.75	38	6.19
	C43 _{P2.3} (P1)	247320	40	1.62	39	5.73
	C43 _{P2.3} (P10)	319696	35	1.09	32	3.65
	C43 _{P2.3} (P60)	300033	58	1.93	38	4.61

		N	Hpl	dHp($\cdot 10^{-4}$)	nM	F _{mín} ($\cdot 10^{-7}$)
Amplicón 1	Q β _{P25}	235849	84	3.56	81	12.63
	C43 _{P25.1} (P1)	228175	78	3.42	75	12.08
	C43 _{P25.1} (P10)	242624	85	3.50	79	11.97
	C43 _{P25.1} (P60)	220119	90	4.09	62	10.35
Amplicón 2	Q β _{P25}	279452	71	2.54	56	6.09
	C43 _{P25.1} (P1)	281442	112	3.98	88	9.50
	C43 _{P25.1} (P10)	263024	76	2.89	57	0.66
	C43 _{P25.1} (P60)	203576	95	4.67	90	13.44
Amplicón 3	Q β _{P25}	351353	75	2.13	74	7.66
	C43 _{P25.1} (P1)	264974	69	2.60	68	9.33
	C43 _{P25.1} (P10)	229209	87	3.80	86	13.64
	C43 _{P25.1} (P60)	236482	85	3.59	78	11.99

Tabla 25: Valores de los índices de diversidad basados en incidencia de las líneas evolutivas C43_{P2.3} y C43_{P25.1} y sus correspondientes ancestros Q β _{P2} y Q β _{P25}. La definición exacta de los parámetros mostrados en la tabla se encuentra en el apartado 4.2.7. de Materiales y Métodos.

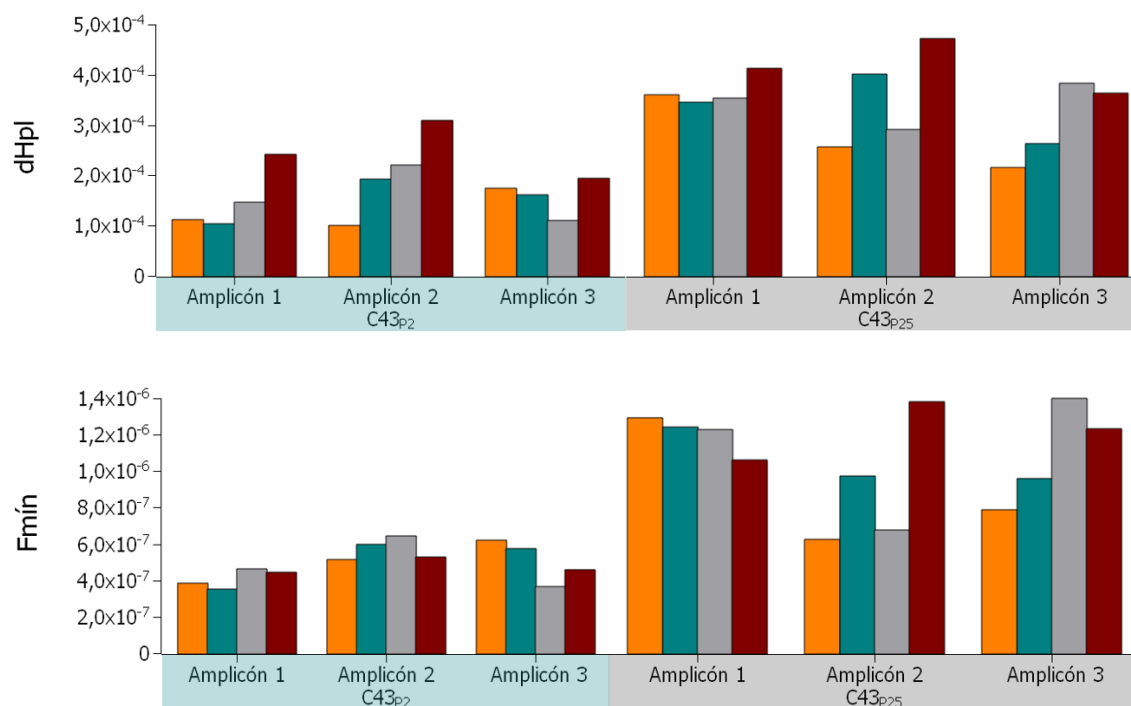


Figura 34: Densidad de haplotipos y frecuencia de mutación mínima a lo largo de la evolución de las líneas C43_{p2}.3 y C43_{p25}.1 Poblaciones ancestrales (●), pase 1 (●), pase 10 (●), pase 60 (●).

Se observa que, en general, la densidad de haplotipos aumenta a lo largo de la adaptación en las dos líneas evolutivas, alcanzándose en todos los casos valores mayores para la línea C43_{p25}.1. En esta línea los mayores aumentos se observan en los amplicones 2 y 3.

En la línea C43_{p2}.3 apenas se observan cambios en la frecuencia de mutación mínima a lo largo de la adaptación. Sin embargo, en los amplicones 2 y 3 correspondientes a la línea C43_{p25}.1 sí se observan aumentos en este parámetro que, a pase 60, presenta valores mayores que para la línea C43_{p2}.3 en todas las regiones. En ambas líneas evolutivas, no se observan grandes diferencias entre regiones en los valores finales alcanzados. Para ambos índices, en la población C43_{p25}.1 se parte de valores mayores que en la C43_{p2}.3 y se llega también a valores mayores a pase 60.

Índices de diversidad basados en abundancia

En la tabla y en las figuras siguientes (Tabla 26 y Figura 35) se muestran los valores de H_s , H_sH , H_sN , $F_{m\acute{a}x}$ y π para las mismas poblaciones, regiones genómicas y números de pase descritos en el apartado anterior.

		N	Hs	HsN($\cdot 10^{-2}$)	HsH	π ($\cdot 10^{-3}$)	$F_{m\acute{a}x}(\cdot 10^{-4})$
Amplicón 1	$Q\beta_{P2}$	266858	0.36	2.88	0.11	0.39	1.98
	C43 _{P2.3} (P1)	270956	0.29	2.36	0.09	0.31	1.55
	C43 _{P2.3} (P10)	239054	0.85	6.86	0.24	1.57	9.14
	C43 _{P2.3} (P60)	265606	1.78	14.22	0.43	3.99	29.46
Amplicón 2	$Q\beta_{P2}$	294855	0.62	4.92	0.16	0.64	3.20
	C43 _{P2.3} (P1)	274730	0.75	5.97	0.19	0.81	4.05
	C43 _{P2.3} (P10)	236250	1.04	8.42	0.26	1.32	6.72
	C43 _{P2.3} (P60)	252206	2.23	17.93	0.51	4.35	27.57
Amplicón 3	$Q\beta_{P2}$	223334	0.40	3.26	0.11	0.40	2.02
	C43 _{P2.3} (P1)	247320	0.42	3.39	0.11	0.43	2.18
	C43 _{P2.3} (P10)	319696	0.67	5.29	0.19	1.07	5.84
	C43 _{P2.3} (P60)	300033	1.76	13.95	0.43	3.59	21.14

		N	Hs	HsN($\cdot 10^{-2}$)	HsH	π ($\cdot 10^{-3}$)	$F_{m\acute{a}x}(\cdot 10^{-4})$
Amplicón 1	$Q\beta_{P25}$	235849	0.91	7.37	0.21	1.04	5.23
	C43 _{P25.1} (P1)	228175	0.89	7.23	0.20	1.05	5.31
	C43 _{P25.1} (P10)	242624	1.17	9.46	0.26	1.70	9.36
	C43 _{P25.1} (P60)	220119	1.76	14.34	0.39	3.15	20.73
Amplicón 2	$Q\beta_{P25}$	279452	1.91	15.23	0.45	3.15	19.11
	C43 _{P25.1} (P1)	281442	2.23	17.74	0.47	3.71	23.36
	C43 _{P25.1} (P10)	263024	1.48	11.89	0.34	2.25	13.93
	C43 _{P25.1} (P60)	203576	1.47	12.00	0.32	1.57	8.13
Amplicón 3	$Q\beta_{P25}$	351353	0.72	5.61	0.17	0.73	3.66
	C43 _{P25.1} (P1)	264974	0.69	5.54	0.16	0.72	3.63
	C43 _{P25.1} (P10)	229209	1.08	8.75	0.24	1.18	5.95
	C43 _{P25.1} (P60)	236482	1.44	11.63	0.32	2.30	11.95

Tabla 26: Valores de los índices de diversidad basados en abundancia de las líneas evolutivas C43_{P2.3} y C43_{P25.1} y sus correspondientes ancestros $Q\beta_{P2}$ y $Q\beta_{P25}$. La definición exacta de los parámetros mostrados en la tabla se encuentra en el apartado 4.2.7. de Materiales y Métodos.



Figura 35: Valores de los índices de diversidad basados en abundancia a lo largo de la evolución de las líneas $C43_{p2.3}$ y $C43_{p25.1}$. Poblaciones ancestrales (●), pase 1 (●), pase 10 (●), pase 60 (●).

Se observa que, en todos los casos analizados, los valores de los índices de diversidad basados en abundancia experimentan grandes ascensos a lo largo de la adaptación a 43 °C, con la única excepción del amplicón 2 en la línea C43_{P25.1}, en el que se producen descensos apreciables. Este resultado es remarcable, ya que las poblaciones ancestrales mostraron los valores más altos para estos mismos parámetros precisamente en el amplicón 2 de Q β _{P25}. En todos los casos, los valores alcanzados por la línea C43_{P2.3} son mayores que los alcanzados por la C43_{P25.1}, lo cual contrasta con el hecho de que la población ancestral Q β _{P2} presentaba valores menores que Q β _{P25} en todas las regiones genómicas analizadas.

Los resultados observados en el amplicón 2 de la línea C43_{P25.1} podrían indicar que la diversidad que se había generado a 37 °C en el ancestro Q β _{P25} no es útil para la adaptación a 37 °C y tiene que ser eliminada, lo cual se traduce en disminuciones en todos los índices de diversidad basados en abundancia. En la línea C43_{P2.3} no se presentaría este problema, ya que su población ancestral ha evolucionado durante menos pases a 37 °C y, por tanto, ha tenido menos tiempo de generar diversidad genética útil a esa temperatura. Otra posible explicación sería que entre los genomas presentes en la población Q β _{P25} ya existiera alguno que aportara alguna ventaja a 43 °C. Este genoma sería favorecido sobre el resto, lo cual tendría un efecto similar a un cuello de botella poblacional, algo que se traduciría también en reducciones de diversidad en etapas tempranas que podrían compensarse o no a lo largo de la adaptación. La pérdida de algunas mutaciones localizadas en este amplicón, que estaban presentes a frecuencias apreciables en la población Q β _{P25} (A2187C, C2201U, G2223A y C2228U), iría a favor de esta hipótesis, sobre todo si se tiene en cuenta que esas mismas mutaciones han sido identificadas en otros experimentos de adaptación a temperatura elevada.

5.3.5 Variación de los índices de diversidad en función del umbral de representación de los haplotipos

Por último, para comprobar si las diferencias encontradas entre las poblaciones C43_{P2.3} y C43_{P25.1} se veían afectadas por el umbral de representación de las secuencias seleccionadas para el análisis, se determinaron los valores de varios índices de diversidad para los umbrales del 0.05 % y 0.1 % en el pase número 60.

Índices de diversidad basados en incidencia

El cálculo de los valores de los índices de diversidad basados en incidencia mostró que estos aumentaban al disminuir el umbral y lo hacían de una forma más marcada en la población C43_{P25.1} que en la C43_{P2.3} (Tabla 27 y Fig. 36). Del mismo modo, las diferencias entre poblaciones también se hicieron mayores al considerar umbrales menores.

		Umbral (%)	N	Hpl	dHpl($\cdot 10^{-4}$)	nM	Fmín($\cdot 10^{-7}$)
Amplificación 1	C43 _{P2.3} (P60)	0.05	265606	64	2.41	32	4.43
		0.10	258268	30	1.16	18	2.56
	C43 _{P25.1} (P60)	0.05	220119	90	4.09	62	1.03
		0.10	208759	30	1.44	22	3.87
Amplificación 2	C43 _{P2.3} (P60)	0.05	252206	78	3.09	44	5.30
		0.10	242939	35	1.44	19	2.38
	C43 _{P25.1} (P60)	0.05	203576	95	4.67	90	13.44
		0.10	194066	41	2.11	40	6.26
Amplificación 3	C43 _{P2.3} (P60)	0.05	300033	58	1.93	38	4.61
		0.10	291787	26	0.89	18	2.24
	C43 _{P25.1} (P60)	0.05	236482	85	3.59	78	11.99
		0.10	228022	42	1.84	39	6.22

Tabla 27: Valores de los índices de diversidad basados en incidencia de las líneas evolutivas C43_{P2.3} y C43_{P25.1} para los umbrales 0.05 % y 0.1 %. La definición exacta de los parámetros mostrados en la tabla se encuentra en el apartado 4.2.7. de Materiales y Métodos.

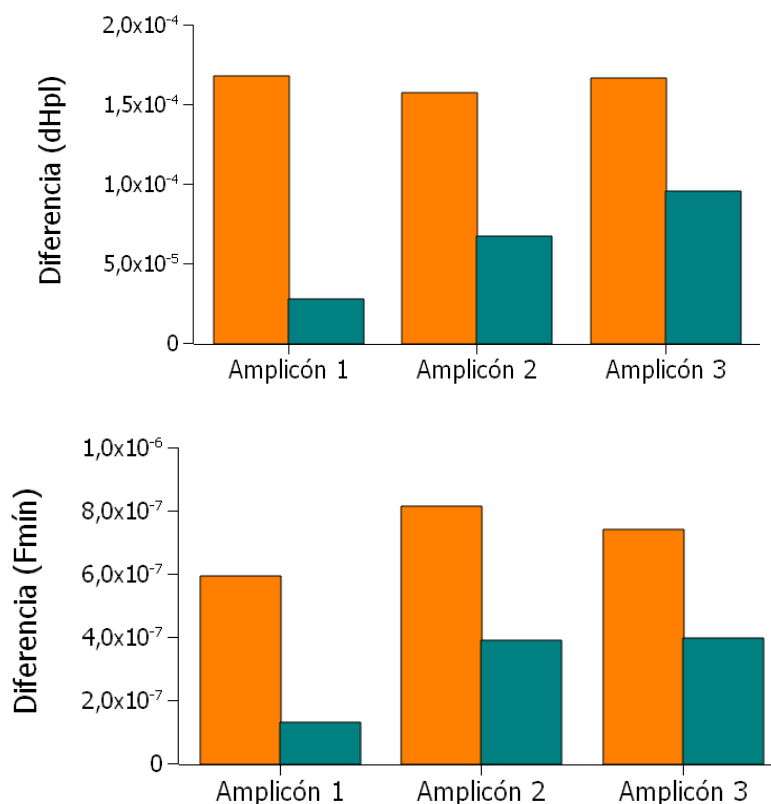


Figura 36: Diferencias entre poblaciones (C43_{P25.1} - C43_{P2.3}) para los valores de los índices de diversidad basados en incidencia. En ambos casos se han utilizados los valores obtenidos en el pase número 60. (●) umbral 0.05 % y (●) umbral 0.1 %. La definición exacta de los parámetros se encuentra en el apartado 4.2.7 de Materiales y Métodos.

Índices de diversidad basados en abundancia

		Umbral (%)	N	Hs	HsN ($\cdot 10^{-2}$)	HsH	π ($\cdot 10^{-3}$)	Fmáx ($\cdot 10^{-4}$)
Amplicón 1	C43 _{P2.3} (P60)	0.05	265606	1.78	14.22	0.43	3.99	29.46
		0.10	258268	1.60	12.81	0.47	3.84	28.72
	C43 _{P25.1} (P60)	0.05	220119	1.76	14.34	0.39	3.15	20.73
		0.10	208759	1.42	11.62	0.42	2.81	19.10
Amplicón 2	C43 _{P2.3} (P60)	0.05	252206	2.23	17.93	0.51	4.35	27.57
		0.10	242939	2.01	16.20	0.56	4.18	26.79
	C43 _{P25.1} (P60)	0.05	203576	1.47	12.00	0.32	1.57	8.13
		0.10	194066	1.15	9.41	0.31	1.32	6.86
Amplicón 3	C43 _{P2.3} (P60)	0.05	300033	1.76	13.95	0.43	3.59	21.14
		0.10	291787	1.58	12.57	0.48	3.43	20.38
	C43 _{P25.1} (P60)	0.05	236482	1.44	11.63	0.32	2.30	11.95
		0.10	228022	1.19	9.67	0.32	2.07	10.82

Tabla 28: Valores de los índices de diversidad basados en abundancia de las líneas evolutivas C43_{P2.3} y C43_{P25.1} para los umbrales 0.05 % y 0.1 %. La definición exacta de los parámetros mostrados en la tabla se encuentra en el apartado 4.2.7. de Materiales y Métodos.

Los índices de diversidad basados en abundancia experimentan cambios menores en función del umbral de representación de las secuencias (Tabla 28 y Fig. 37), lo cual hace que las relaciones entre poblaciones se mantengan con pocas variaciones.

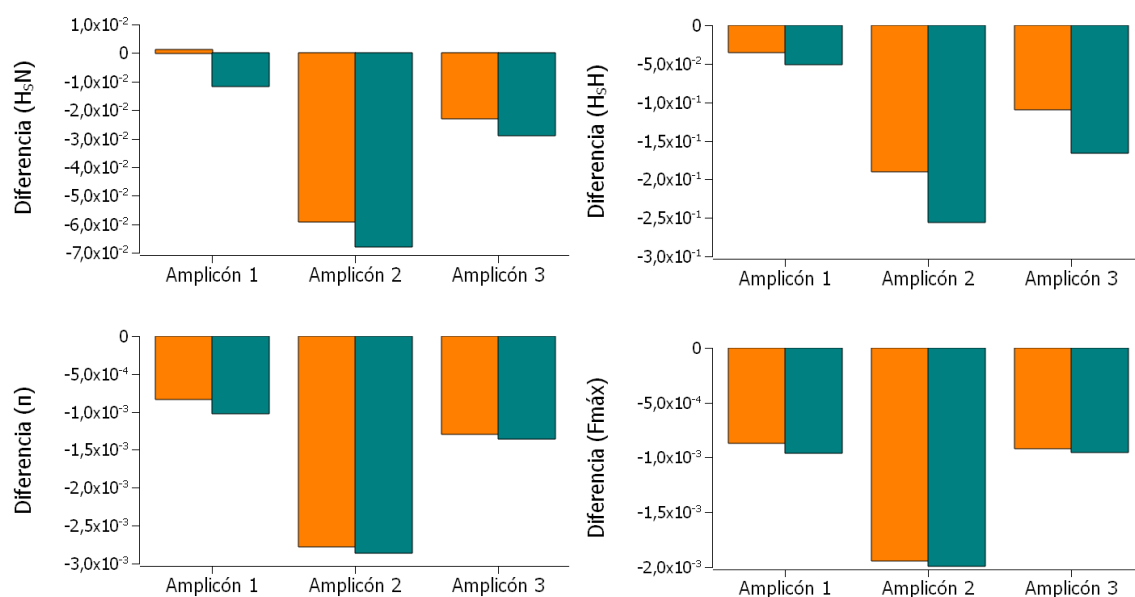


Figura 37: Diferencias entre poblaciones (C43_{P25.1} - C43_{P2.3}) para los valores de los índices de diversidad basados en abundancia. En ambos casos se han utilizado los valores obtenidos en el pase número 60. (●) umbral 0.05 % y (●) umbral 0.1 %. La definición exacta de los parámetros se encuentra en el apartado 4.2.7 de Materiales y Métodos.

5.4 CAPÍTULO 4

Influencia de la pauta de cambio en la adaptación a temperatura elevada

Otro de los factores que puede influir en la adaptación a una nueva presión selectiva es la velocidad a la que esta varía. Para estudiar la importancia de este parámetro en la adaptación a temperatura elevada del bacteriófago Q β , en primer lugar se evaluó su capacidad replicativa a varias temperaturas comprendidas entre 37 °C y 43 °C (Fig. 38). Los resultados obtenidos mostraron un claro descenso de los títulos virales en función de la temperatura.

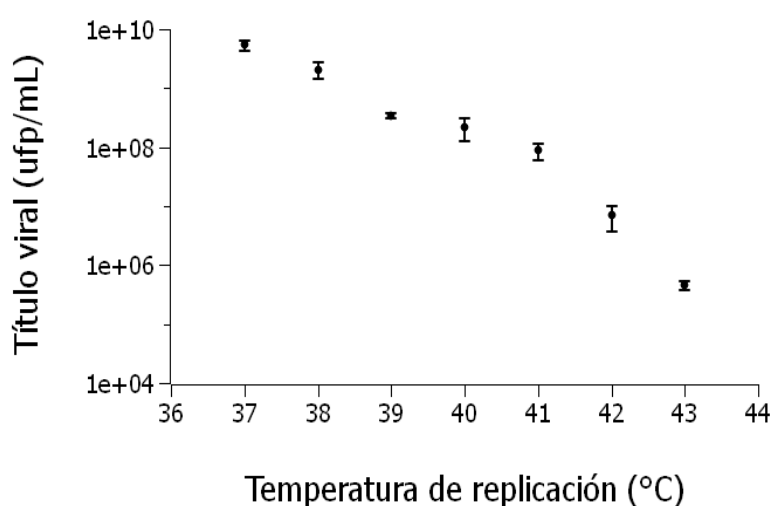


Figura 38: Capacidad replicativa del bacteriófago Q β a distintas temperaturas. El ensayo se realizó incubando a cada temperatura 10^6 ufp de la población Q β_{P25} con aproximadamente 10^8 bacterias (*E. coli* Hfr) en fase exponencial, en un volumen de 1 mL. Las condiciones de la incubación y los métodos para estimar la producción viral están descritos en Materiales y Métodos (apartado 4.2.2.).

A continuación, la población ancestral Q β_{P25} (ver Capítulo 1) se propagó a temperaturas más altas de la óptima siguiendo las pautas de cambio C43, G1 y G2 descritas en el apartado 4.2.2. de Materiales y Métodos (Fig. 10), hasta llegar al valor máximo de 43 °C. Para evaluar el efecto de la propagación constante a 43 °C desde el primer pase se utilizaron las líneas C43 $_{P25}$ descritas en los Capítulos 2 y 3. Los otros dos patrones de cambio implicaron aumentos de temperatura progresivos, en los que la temperatura de 43 °C se alcanzó en dos etapas (30 pases a 40 °C seguidos de 30 pases a 43 °C; líneas G1 $_{P25}$) o en seis (aumentos de 1 °C cada 10 pases; líneas G2 $_{P25}$). Todas las propagaciones se realizaron por triplicado para un total de 60 pases a temperatura superior a la óptima. Es importante tener en cuenta que el número de pases experimentados a 43 °C dependió del patrón de cambio ensayado: 60, 30 o 10 pases para las líneas evolutivas C43 $_{P25}$, G1 $_{P25}$ y G2 $_{P25}$ respectivamente.

5.4.1 Variación del *fitness* a 43 °C a lo largo de la adaptación

Para analizar la dinámica de la adaptación de Q β al aumento de temperatura en función de la velocidad de cambio en este parámetro, se determinaron los valores de *fitness* a 43 °C cada 10 pases, para las líneas evolutivas propagadas según los patrones de cambio descritos (Fig. 39). Los valores obtenidos se refirieron al valor de *fitness* de la población ancestral Q β _{P25} que, como se indicó en el Capítulo 1, fue de 4.8 ± 0.7 .

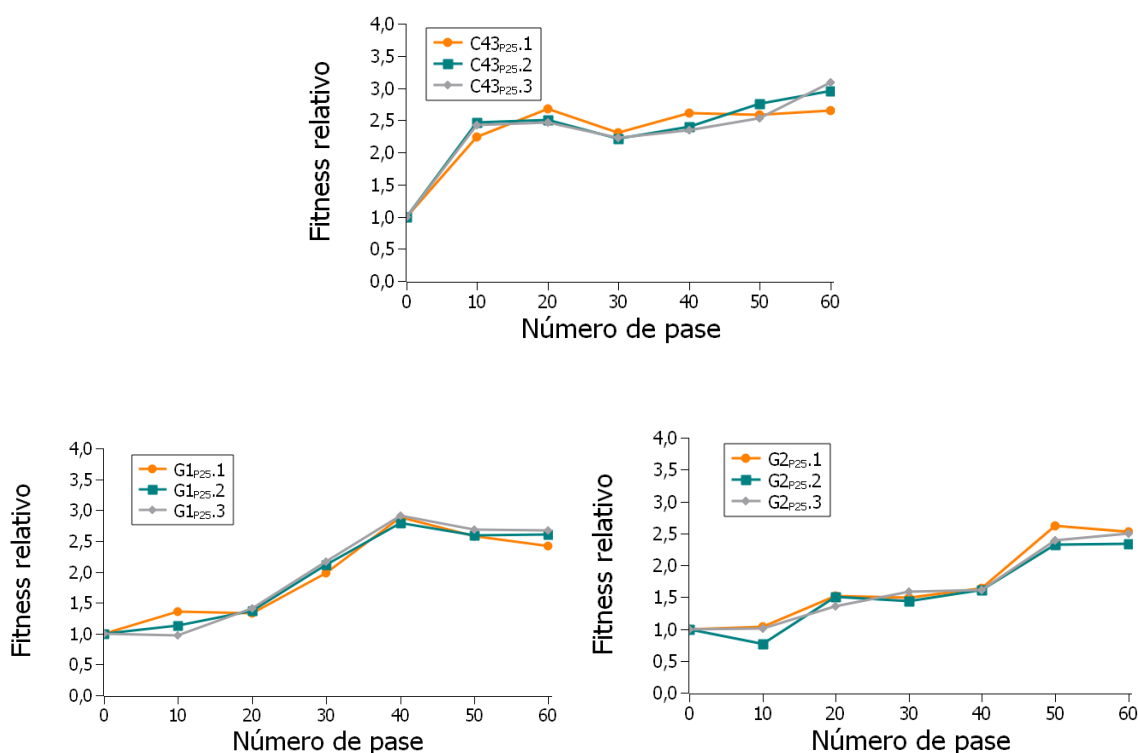


Figura 39: Dinámica de la variación del *fitness* a 43 °C en las líneas evolutivas C43P25, G1P25 y G2P25.

Los resultados obtenidos muestran una notable influencia de la velocidad de aumento de la temperatura en la dinámica de ganancia de *fitness* a 43 °C. Los aumentos más rápidos se produjeron en las líneas C43P25, seguidas de las G1P25. En las líneas G2P25 se observó que, aunque la propagación del virus a temperaturas superiores a 38 °C ya produjo aumentos de *fitness* a 43 °C, estos fueron muy leves hasta que no se alcanzó la temperatura de 42 °C. Los mayores aumentos de *fitness* tuvieron lugar entre los pases 1 y 10 en las líneas C43P25, entre los pases 20 y 40 en las líneas G1P25 (los últimos 10 pases a 40 °C y los primeros 10 a 43 °C) y, por último, entre los pases 40 y 50 (42 °C) en las líneas G2P25.

Para hacer comparaciones cuantitativas entre las diferentes líneas evolutivas, se determinaron el número de pases necesarios para duplicar el valor de *fitness* de la población ancestral (es decir, para alcanzar un valor de *fitness* relativo de 2), el *fitness* alcanzado en el último pase a 43 °C (pase 60) y las diferencias durante los primeros 10 pases a 43 °C. Estos valores se muestran en la Tabla 29.

Línea evolutiva	Número de pase (fitness relativo = 2)	Fitness relativo a pase 60	Diferencia de fitness (primeros 10 pases a 43 °C)
C43_{P25}.1	8.1	2.7	1.2
C43_{P25}.2	6.8	3.0	1.5
C43_{P25}.3	7.0	3.1	1.4
C43_{P25} (Media ± SD)	7.3 ± 0.7	2.9 ± 0.2	1.4 ± 0.1
G1_{P25}.1	30.3	2.4	0.9
G1_{P25}.2	28.5	2.6	0.7
G1_{P25}.3	27.8	2.7	0.8
G1_{P25} (Media ± SD)	28.9 ± 1.3	2.6 ± 0.1	0.8 ± 0.1
G2_{P25}.1	43.7	2.5	-0.1
G2_{P25}.2	45.5	2.3	0.0
G2_{P25}.3	45.0	2.5	0.1
G2_{P25} (Media ± SD)	44.7 ± 0.9	2.5 ± 0.1	0.0 ± 0.1

Tabla 29: Parámetros cuantitativos que describen la dinámica de la variación del *fitness* a 43 °C en las líneas evolutivas C43_{P25}, G1_{P25} y G2_{P25}. Media y SD hacen referencia a los valores medios y su correspondiente desviación estándar.

Las poblaciones que evolucionaron a través de aumentos de temperatura progresivos (G1_{P25} y G2_{P25}) alcanzaron valores de *fitness* estadísticamente similares en el pase número 60 ($p > 0.1$ para la comparación de las líneas G1_{P25} y G2_{P25}, test *t*-Student) que, sin embargo, fueron ligeramente inferiores a los alcanzados por las líneas C43_{P25} en el mismo pase ($p < 0.005$ para las comparaciones de las líneas C43_{P25} con G1_{P25} y G2_{P25}, test *t*-Student). También hubo diferencias importantes en la cantidad de pases requeridos para duplicar el valor de *fitness* inicial, que fue mayor cuanto más lento fue el aumento de temperatura, ($p < 0.001$ para todas las posibles comparaciones por pares entre los diferentes conjuntos de líneas, test *t*-Student). Por último, La variación de *fitness* en los 10 primeros pases experimentados a 43 °C (esto es del 1 al 10 en C43_{P25}, del 30 al 40 en G1_{P25} y del 50 al 60 en G2_{P25}) fue mayor cuanto más brusco fue el cambio de temperatura (C43_{P25} > G1_{P25} > G2_{P25}), observándose una relación inversa entre la ganancia de *fitness* en esos pases y el valor de *fitness* en el pase anterior a la primera exposición a 43 °C.

5.4.2 Dinámica de adquisición de mutaciones en la secuencia consenso

Con el objetivo de analizar la dinámica de la adaptación a nivel genético, se determinó la secuencia consenso de las diferentes líneas evolutivas cada 10 pases (Tabla 17.b para C43_{P25} y Tabla 30 para G1_{P25} y G2_{P25}). Lo primero que se observa es que las mutaciones C3065U y A2187C, que estaban presentes como polimorfismos en el ancestro Qβ_{P25}, mostraron un comportamiento diferente durante la evolución a temperatura elevada. Mientras que C3065U se perdió en todas las líneas evolutivas, A2187C presentó un comportamiento no uniforme, desapareciendo en todas las líneas C43_{P25} y fijándose en todas las G1_{P25} y G2_{P25}, con una única excepción en la que se mantuvo como polimorfismo. Es destacable que esta mutación también se mantuvo como polimorfismo, a pase número 60, en todas las líneas C43_{P2}.

Tabla 30.a: Líneas evolutivas G1_{P25}

Tabla 30.a: Líneas evolutivas G1P25

	wt	mut<wt	mut=wt	mut>wt	mut									
	A1088G	U1295C/G	G1312A	G1371A	A1778G	C1806U	A1930G	A2187C	C2201U	G2223A	C2384U	U2776C	C3545U	G3945A
P10														
P20														
P30														
P40														
P50														
P60														

Tabla 30.b: Líneas evolutivas G2_{P25}

	G502C	A1088G	U1295C/G	G1312A	G1371A	A1778G	C1806U	A1930G	A2187C	C2201U	G2223A	U2776C	C3545U	G3945A
P10														
P20														
P30														
P40														
P50														
P60														

Tabla 30: Mutaciones y polimorfismos en las secuencias consenso de las líneas evolutivas G1_{P25} y G2_{P25}.

La primera fila muestra los diferentes cambios nucleotídicos observados. En las filas siguientes aparece el número de pase analizado (P10 a P60). El estado de las mutaciones se muestra con un sombreado de diferente intensidad (ver parte superior de la tabla) en los cuadrados situados debajo de cada mutación (de izquierda a derecha, réplicas 1, 2 y 3 respectivamente). La posición 1295 experimentó cambios U/C o U/G que, con frecuencia, coexistían en la misma población. En la figura no se ha hecho distinción entre ambos tipos de cambio.

En general se observa que la mayoría de las mutaciones incrementan su representación a lo largo de la serie de pases. Sin embargo, también hay algunas que aparecieron como polimorfismos en algún punto y desaparecieron o disminuyeron sus frecuencias a pases posteriores. En algunos casos esas mismas mutaciones volvieron a reaparecer más adelante. El ejemplo más notable fue la sustitución A1930G, presente en el pase 10 de todas las líneas G1_{P25} y G2_{P25} y que llegó a ser indetectable en la secuencia consenso de la mayoría de las líneas cuando estas fueron propagadas a temperaturas más próximas a 43 °C.

La representación del número total de mutaciones (fijadas y polimórficas) adquiridas por cada línea evolutiva en función del número de pase (Fig. 40) permitió analizar si la dinámica de adquisición de mutaciones es similar a la seguida por las variaciones de *fitness* (Fig. 39) y si hay diferencias en función de la pauta de cambio de temperatura.

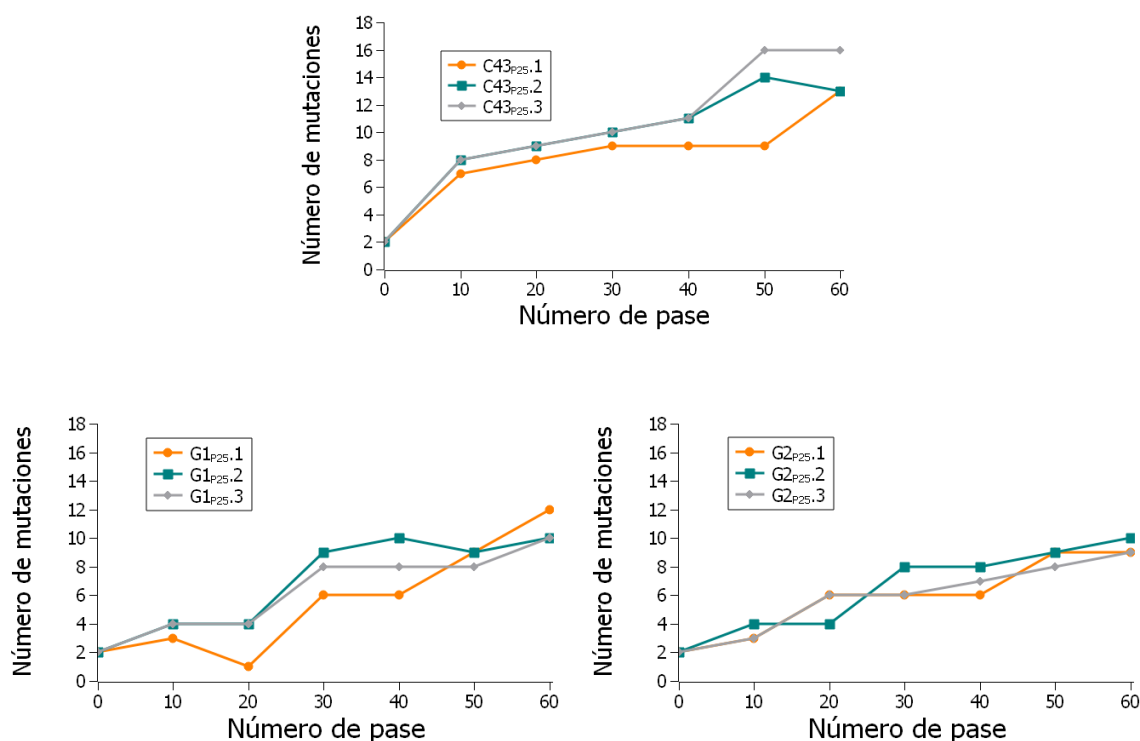


Figura 40: Número total de mutaciones (fijadas y polimórficas) en las secuencias consenso de las líneas C43P25, G1P25 y G2P25 en función del número de pase.

En general, se observa que las variaciones del número de mutaciones en función del número de pase siguen tendencias similares a las experimentadas por los valores de *fitness* (Fig. 39). Durante los primeros 10 pases, las líneas C43P25 adquirieron mutaciones más rápidamente que el resto, lo que concuerda con su adaptación más rápida en las primeras etapas.

La distribuciones finales (a pase número 60) de las mutaciones en el genoma de las poblaciones G1P25 y G2P25 se muestran en la Fig. 41. La distribución correspondiente a la población C43P25 se puede ver en la Fig. 27 (Capítulo 2 apartado 5.2.2.).

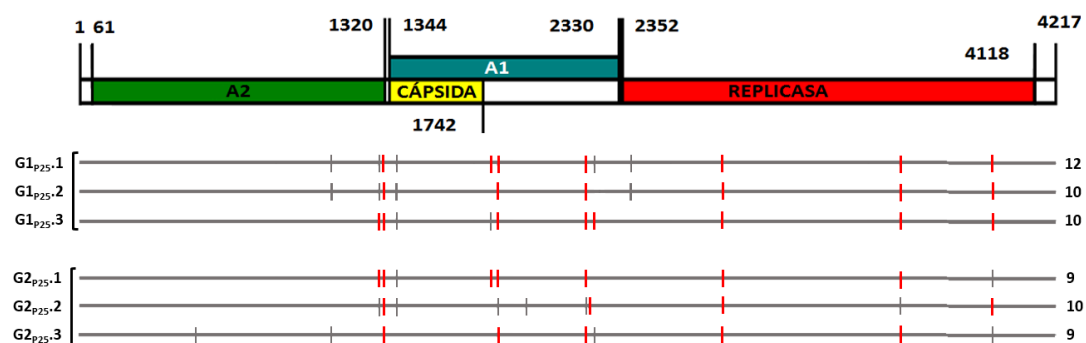


Figura 41: Mutaciones en las secuencias consenso de las líneas G1_{P25} y G2_{P25} en el pase número 60 de la adaptación a 43 °C. Las mutaciones fijadas en las poblaciones se representan con una línea roja y las mutaciones en estado polimórfico con una línea gris. A la derecha se indica el número de mutaciones totales en cada una de las líneas. La región secuenciada es la comprendida entre los nucleótidos 280 y 4150, según se indica en el apartado 4.2.5.

Al igual que sucedía en el caso de las líneas C43_{P25} y C43_{P2} (Fig. 27), en este caso también se observa que la distribución de las mutaciones no es uniforme. Hay regiones en las que hay una mayor concentración de mutaciones (localizadas en los genes de la proteína mayoritaria de la cápsida y A1), y otras con densidad menor (Gran parte del gen de la proteína A2 y algunas regiones de la replicasa).

5.4.3 Comparación de las mutaciones adquiridas en la secuencia consenso en función de la pauta de cambio de temperatura

Las líneas evolutivas obtenidas a pase 60, mostraron un total de 33 mutaciones que afectaron 32 posiciones nucleotídicas (Tabla 31). Aunque la posición 1295 puede experimentar un cambio U/C o U/G, que implica una sustitución de fenilalanina por serina o cisteína respectivamente, al igual que se hizo en el Capítulo 2 cuando se compararon las poblaciones C43_{P2} y C43_{P25}, se considerarán ambos cambios como una única mutación. Once de los cambios nucleotídicos observados fueron sinónimos y 21 implicaron un cambio de aminoácido. Las líneas C43_{P25} presentaron un total de 23 mutaciones diferentes, un número que se redujo a 12 en el caso de las G1_{P25} y a 14 para las G2_{P25}. A pesar de que todas las líneas analizadas en esta sección procedían de la misma población ancestral, tan solo 6 de las 32 mutaciones coincidieron en alguna de las réplicas de las 3 pautas de cambio. El número de mutaciones comunes aumentó a 11 cuando solo se consideraron los tratamientos progresivos, G1_{P25} y G2_{P25}, mientras que las mutaciones que aparecen de forma exclusiva en los linajes C43_{P25}, G1_{P25} y G2_{P25} fueron 17, 1 y 3 respectivamente (Fig. 42).

Mutación ¹	Gen ²	Número de líneas con la mutación ³		
		C43 _{P25}	G1 _{P25}	G2 _{P25}
G502C (G147R)	A2			0/1
A1088G (D342G)	A2	3/0	0/2	0/1
C1228U (L389F)	A2	1/0		
U1295(C/G) (F411S/C)	A2	2/1	1/2	1/1
G1312A (V417I)	A2	3/0	3/0	3/0
G1371A (G9S)	C	3/0	0/3	0/2
U1544C (S)	C	1/0		
G1607U (Q87H)	C	1/0		
C1649U (S)	C	2/0		
A1778G (S)	A1		1/1	1/0
C1806U (P154S)	A1		3/0	2/1
A1827G (T161A)	A1	1/0		
A1930G (Q195R)	A1			0/1
A1956G (K204E)	A1	0/1		
U2006G (S220R)	A1	0/1		
U2016C (F224L)	A1	2/0		
U2135C (S)	A1	1/0		
A2187C (S281R)	A1		3/0	2/1
C2201U (S)	A1			1/0
A2222C (S)	A1	3/0		
G2223A (V293I)	A1		1/1	0/1
C2384U (S)	R		0/2	
U2641G (L96R)	R	1/0		
U2776C (V141A)	R	1/1	3/0	3/0
G3002A (S)	R	1/0		
U3011C (S)	R	1/0		
U3311G (I320M)	R	3/0		
G3521U (E389D)	R	1/0		
C3545U (S)	R		3/0	2/1
G3896A (S)	R	0/1		
C3903U (L517F)	R	2/1		
G3945A (G531S)	R	2/1	3/0	1/2

Tabla 31: Mutaciones presentes en las líneas evolutivas C43_{P25}, G1_{P25} y G2_{P25} en el pase número 60 de la adaptación a 43 °C. ¹Todos los cambios se identificaron respecto a la secuencia del virus Qβ_{WT}. El cambio de aminoácido se muestra entre paréntesis. S indica que la mutación es sinónima. La posición 1295 puede mutar a C o G. ²Ubicación de la mutación en los diferentes genes del genoma de Qβ (C: cápsida, R: replicasa). ³El primer número de cada par indica el número de líneas en un conjunto dado en el que se fija la mutación. El segundo número se refiere al número de líneas en las que la mutación está presente como polimorfismo. Las celdas vacías indican la ausencia de la mutación en las líneas de ese conjunto en particular.

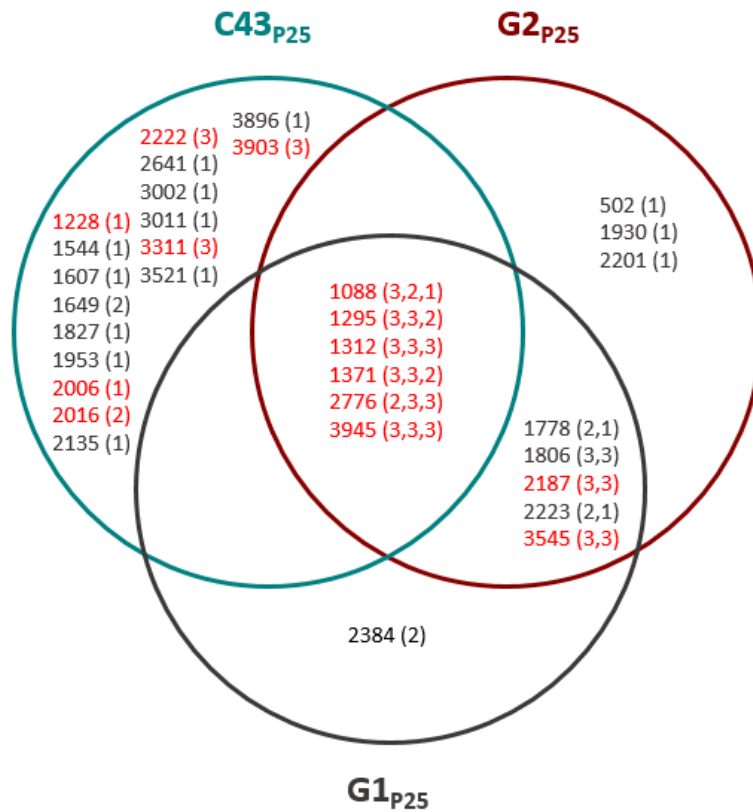


Figura 42: Diagrama de Venn mostrando las mutaciones coincidentes y exclusivas de las líneas evolutivas C43_{p25}, G1_{p25} y G2_{p25}. La circunferencia correspondiente a cada conjunto de líneas contiene las posiciones nucleotídicas que aparecen mutadas a pase 60 en al menos una línea. Entre paréntesis se indica el número de réplicas en las que la mutación está presente. En el caso de mutaciones coincidentes en los tres conjuntos, el primero de los números del paréntesis se refiere a las réplicas de C43_{p25}, el segundo a las de G1_{p25} y el tercero a las de G2_{p25}. Para las mutaciones coincidentes en G1_{p25} y G2_{p25}, el primer número se refiere a las de G1_{p25} y el segundo a las de G2_{p25}. En rojo se denotan las mutaciones que también se encontraban en C43_{p2}.

Es destacable que las 6 mutaciones comunes a todas las pautas de cambio son no sinónimas y están altamente representadas (al menos en 6 de las 9 líneas analizadas). Esto contrasta con las mutaciones exclusivas de cada conjunto de líneas, las cuales en muchos casos están representadas en una única línea e incluyen mutaciones sinónimas. Mientras que las líneas C43_{p25} no presentan ninguna mutación en común con solo uno de los conjuntos G1_{p25} o G2_{p25}, estos últimos sí presentan varias mutaciones comunes entre ellos.

A continuación se determinó cuáles fueron las mutaciones que primero se fijaron para cada una de las poblaciones en estudio y el pase en el que tuvo lugar dicha fijación.

Línea evolucionada	Primeras mutaciones fijadas	Número de pase
C43_{P25.1}	A1088G, C1649U, U2016C, A2222C, U2460C, U3311G, G3945A	10
C43_{P25.2}	A1088G, A2222C, U3311G	10
C43_{P25.3}	A1088G, C1649U, A2222C, U3311G	10
G1_{P25.1}	G1312A	20
G1_{P25.2}	G1312A, A1930G	20
G1_{P25.3}	G1312A	30
G2_{P25.1}	G1312A	20
G2_{P25.2}	A1930G	20
G2_{P25.3}	G1312A	20

Tabla 32: Primeras mutaciones fijadas durante la evolución a 43 °C de las líneas C43_{P25}, G1_{P25} y G2_{P25}. En la tercera columna de la tabla, se indica el número de pase en el que se detectó el primer evento de fijación de mutaciones.

Se observa que, en las poblaciones que evolucionaron de forma constante a 43 °C, el primer evento de fijación de mutaciones tuvo lugar en un pase más temprano e involucró un número mayor de mutaciones que en las poblaciones que alcanzaron los 43 °C de forma progresiva. En las líneas G1_{P25} y G2_{P25}, la primera mutación fijada fue G1312A, con la única excepción de G2_{P25.2}, en la que primero se fijó A1930G. Solo en un caso (G1_{P25.2}) se fijaron dos mutaciones de manera simultánea (A1930G y G1312A). En ningún caso estas sustituciones se encontraron entre las primeras fijadas en las líneas C43_{P25}, aunque G1312A sí que pudo ser detectada a pases posteriores. La mutación A1930G no fue detectable en ninguno de los pases de estas líneas y se perdió o disminuyó su intensidad en el resto. También es remarcable que la mayoría de las sustituciones que están entre las primeras fijadas en las líneas C43_{P25} (C1649U, U2016C, A2222C, U2460C y U3311G) no fueron detectadas en la secuencia consenso de ninguno de los pases analizados en G1_{P25} y G2_{P25} o solo lo hicieron a pases tardíos como polimorfismos (A1088G).

5.4.4 Grado de convergencia durante la adaptación

Para determinar de forma cuantitativa el grado de similitud entre las vías evolutivas seguidas por las poblaciones evolucionadas a través de la misma pauta de cambio, se analizó la diversidad nucleotídica (ver apartado 5.2.4.) entre sus secuencias consenso obtenidas a pase 60. La evolución a través de cualquiera de los tratamientos progresivos condujo a una menor diversidad de nucleótidos (6.89×10^{-4} para G1_{P25} y 1.38×10^{-3} para G2_{P25}) que la observada en las líneas evolucionadas de forma constante a 43 °C, C43_{P25} (2.58×10^{-3}). La determinación de los índices de convergencia, definidos en la Fig. 29 del Capítulo 2, mostró que los valores mayores se obtenían para las líneas G1_{P25} (Tabla 33).

	C43 _{P25}	G1 _{P25}	G2 _{P25}
<i>iC₁</i>	0.35	0.67	0.43
<i>iC₂</i>	0.48	1.00	0.57
<i>iC₃</i>	0.25	0.67	0.22
<i>iC₄</i>	0.50	0.67	0.56

Tabla 33: Índices de convergencia para las poblaciones C43_{P25}, G1_{P25} y G2_{P25} en el pase 60 de la adaptación a 43 °C. Cada uno de estos valores ha sido calculado como se indica en la Fig. 29.

El cálculo de la convergencia evolutiva entre diferentes conjuntos de líneas (calculado como el número de mutaciones presentes en al menos una línea de cada conjunto dividido entre el número total de mutaciones en los conjuntos considerados) dio un valor de 0.73 para G1_{P25} y G2_{P25}. Ese valor se redujo notablemente cuando las líneas C43_{P25} se compararon con cada uno de los conjuntos G1_{P25} o G2_{P25}, obteniéndose valores de 0.21 y 0.19 respectivamente. La comparación de las tres líneas, C43_{P25}, G1_{P25} y G2_{P25} mostró un valor de 0.19.

5.5 CAPÍTULO 5

Estudio mediante NGS de la variación de la diversidad genética intrapoblacional en función de la pauta de cambio en las presiones selectivas

Al igual que se hizo para las poblaciones que evolucionaron de forma constante a 43 °C, se realizó un análisis mediante NGS de una línea representativa de cada conjunto de líneas G1_{P25} (línea 2) y G2_{P25} (línea 1) que alcanzaron la temperatura de 43 °C de forma gradual. Las regiones genómicas secuenciadas, así como los métodos de análisis fueron los mismos ya descritos para las líneas Q β _{P25} y Q β _{P2}. Los pases analizados fueron el pase número 60, el último realizado a una temperatura inferior a 43 °C (30 en G1_{P25} y 50 en G2_{P25}) y el primero y el décimo realizados a 43 °C (pases 31 y 40 en G1_{P25} y 51 y 60 en G2_{P25}). En todos los casos, los resultados obtenidos se compararon con los de la línea C43_{P25.1}, que también fue analizada en el Capítulo 3. El número de secuencias y haplotipos totales obtenidos, así como los resultantes después de quedarnos con los que están representados con una frecuencia mayor del 0.05 % se muestra en la Tabla 6.

5.5.1 Distribuciones de haplotipos

Las distribuciones de frecuencias de los haplotipos representados por encima del 0.05 % en el pase número 60 de las tres líneas analizadas se muestra en la Fig. 43. En general, no se observan diferencias que puedan ser atribuidas a la pauta de cambio, aunque parece que en los amplicones 1 y 2 los haplotipos que tienen mayores frecuencias están más representados en las poblaciones graduales que en la evolucionada de forma constante.

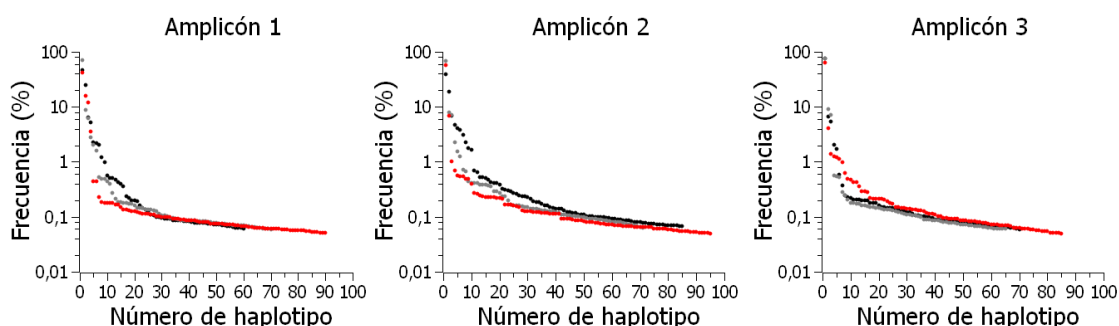


Figura 43: Frecuencias de los haplotipos representados en más del 0.05 % de las lecturas obtenidas a pase número 60 en cada uno de los amplicones analizados. C43_{P25} (●), G1_{P25.2} (●) y G2_{P25.1} (●). El número de cada haplotipo representa el número de orden en relación a su frecuencia.

Con objeto de evaluar los cambios que se producen durante los diez primeros pases a 43 °C, se compararon las distribuciones de haplotipos obtenidas en ese punto de la evolución con las obtenidas en el pase previo al primero realizado a 43 °C (Fig. 44).

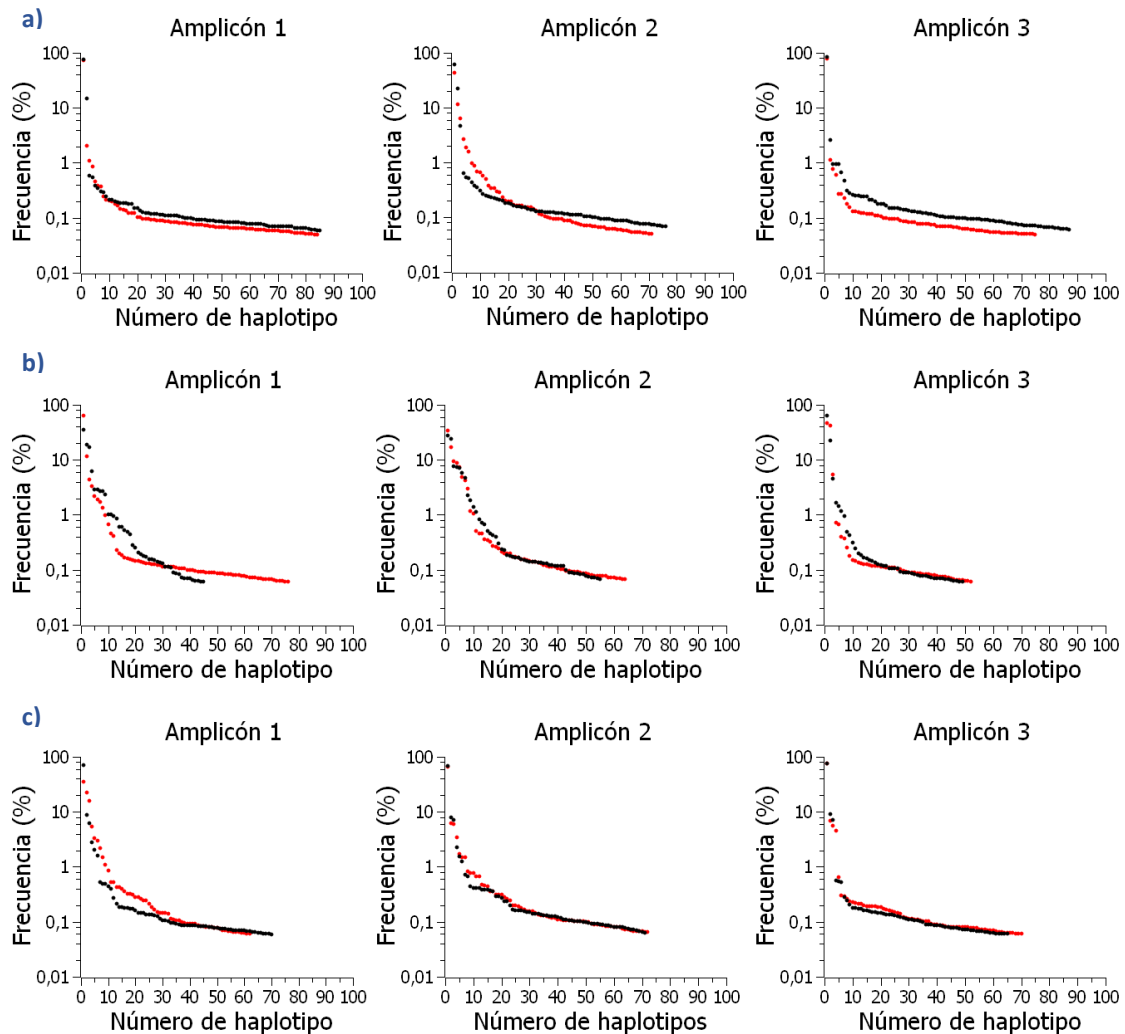


Figura 44: Cambios en la distribución de frecuencias de haplotipos durante los 10 primeros pases a 43 °C en cada uno de los amplicones analizados. a) Q β _{P25} (●) y C43_{P25.1} pase 10 (●), b) G1_{P25.2} pase 30 (●) y pase 40 (●) y c) G2_{P25.1} pase 50 (●) y pase 60 (●). El número de cada haplotipo representa el número de orden en relación a su frecuencia. Los haplotipos incluidos en la representación son los que están representados en más del 0.05 % de las lecturas en cada amplicón y pase analizados.

Aunque no se observa un patrón claro que diferencie a las tres poblaciones, si nos centramos en el amplicón 2, se ve que en la población C43_{P25.1}, durante los primeros 10 pases a 43 °C se produce una clara disminución de las frecuencias de los haplotipos más representados (con la única excepción del primero). Este resultado es coherente con los descensos de diversidad que se observaban en esta región genómica durante la adaptación a 43 °C.

La distribución de las frecuencias de haplotipos en función de su distancia de Hamming a la secuencia consenso se muestra en la Fig. 45.

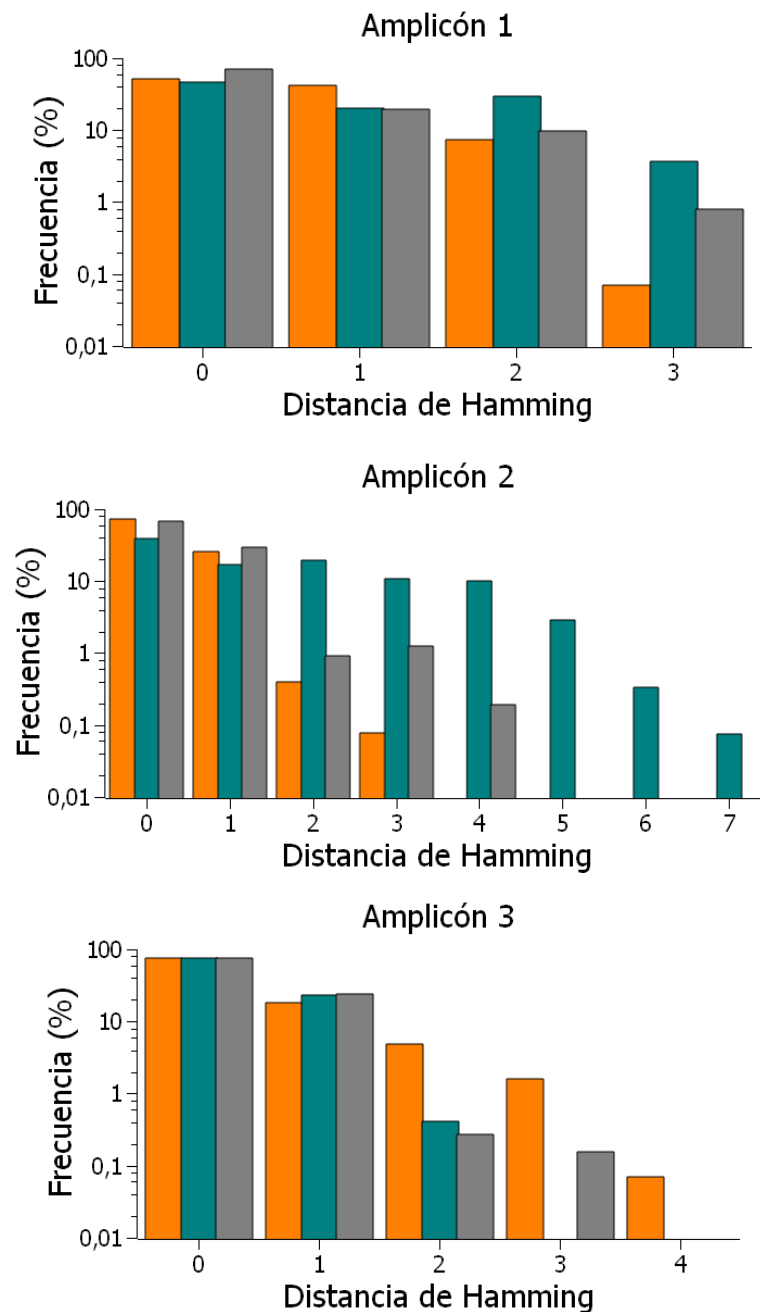


Figura 45: Frecuencia de haplotipos en función de su distancia de Hamming respecto a la secuencia consenso. Poblaciones C43P25.1 (●), G1P25.2 (●) y G2P25.1 (●) obtenidas a pase número 60.

Nuevamente las diferencias más claras se encuentran en el amplicón 2 en el que puede observarse que las líneas G1P25.2 y G2P25.1 poseen distribuciones más amplias de haplotipos, en las que las frecuencias de los mismos en función de su distancia de Hamming decaen más lentamente que en C43P25.1.

5.5.2 Evolución de la frecuencia de las mutaciones más representadas a lo largo de la adaptación

A continuación, se realizó un seguimiento de las mutaciones que ya estaban presentes en las poblaciones ancestrales en una proporción mayor del 0.5 % (Tablas 34 y 35) y de las que, sin haber sido identificadas en la población ancestral, alcanzaron una proporción por encima del 5 % en alguno de los pases a temperaturas mayores de 37 °C (Tabla 36).

		Frecuencia (%)				
	Mutación	Q β _{P25}	G1 _{P25.2} (P30)	G1 _{P25.2} (P31)	G1 _{P25.2} (P40)	G1 _{P25.2} (P60)
Amplificón 1	A1088G	0.52	3.66	4.61	48.18	61.00
	U1295C	1.35	17.16	15.88	7.60	3.57
	U1295G	0.52	2.68	3.72	27.33	34.41
	G1312A	2.85	82.43	83.79	88.98	96.89
	U1328G	0.99	0.82	0.90	0.55	0.70
Amplificón 2	A2187C	68.80	47.04	52.84	55.25	87.00
	C2189U	1.13	0.00	0.00	0.00	0.00
	C2201U	6.21	55.98	45.09	35.60	16.70
	G2217A	2.23	0.58	0.64	0.51	1.53
	G2223A	2.93	5.95	5.40	3.11	2.65
	U2225C	1.69	0.24	0.36	0.00	0.28
	C2228U	8.62	10.19	9.74	8.15	4.78
	G2253A	1.11	0.00	0.00	0.00	0.00
	U2255C	0.87	0.00	0.08	5.88	3.53
Amplificón 3	U3402C	0.90	6.10	6.55	2.12	5.44
	C3545U	1.29	45.10	48.60	73.72	92.95
	A3603G	0.71	1.03	0.58	0.63	0.65

Tabla 34: Trayectoria seguida por las mutaciones representadas en la población ancestral en proporción mayor o igual al 0.5 % durante la evolución de la población G1_{P25.2}. Las celdas sombreadas en gris indican aumentos en la frecuencia de la mutación indicada (gris claro entre la población ancestral y el pase 30 y gris oscuro entre los pases 31 y 60), mientras que las sombreadas en rosa indican descensos de frecuencia (rosa claro entre la población ancestral y el pase 30 y rosa oscuro entre los pases 31 y 60). Las celdas correspondientes a mutaciones que no han alcanzado una representación del 5 % en ningún pase, o lo han hecho solo en uno, no se han resaltado.

En la línea G1_{P25.2}, de las 17 mutaciones que estaban presentes en la población ancestral, 9 experimentan ascensos en sus frecuencias durante los 30 primeros pases realizados a 40 °C (A1088G, U1295C, U1295G, G1312A, C2201U, G2223A, C2228U, U3402C, C3545U). La única mutación que, estando representada en una proporción mayor del 5 % en la población ancestral, desciende su frecuencia durante estos pases es

A2187C, la cual muestra un comportamiento similar, aunque mucho menos acusado, al mostrado durante la evolución constante a 43 °C (ver Tabla 21 en el Capítulo 3). Durante los siguientes 30 pases a 43 °C, algunas de las mutaciones que habían aumentado sus frecuencias durante la propagación a 40 °C (U1295C, C2201U, G2223A, C2228U y U3402C) experimentan un descenso en su representación. El resto sigue la tendencia creciente. Por el contrario, la mutación A2187C que había disminuido su representación durante la propagación a 40 °C experimenta un notable ascenso durante su propagación a 43 °C entre los pases 30 y 60, algo que contrasta con su total desaparición durante la evolución constante a 43 °C (ver Tabla 21 en el Capítulo 3). El resto de mutaciones presentes en la población ancestral experimentan pocos cambios, aunque en general tienen una tendencia descendente. A pase número 60, la línea G1_{P25.2} presenta 7 mutaciones de la población ancestral con frecuencias por encima del 5 %.

	Mutación	Frecuencia (%)			
		Qβ _{P25}	G2 _{P25.1} (P50)	G2 _{P25.1} (P51)	G2 _{P25.1} (P60)
Amplicón 1	A1088G	0.52	7.00	8.16	9.49
	U1295C	1.35	23.91	24.18	10.42
	U1295G	0.52	24.43	33.43	77.85
	G1312A	2.85	97.04	97.27	99.38
	U1328G	0.99	0.92	0.63	0.45
Amplicón 2	A2187C	68.80	91.48	93.30	99.22
	C2189U	1.13	0.00	0.00	0.00
	C2201U	6.21	8.48	8.75	8.23
	G2217A	2.23	2.00	1.58	0.42
	G2223A	2.93	3.89	4.35	7.62
	U2225C	1.69	0.78	0.71	0.38
	C2228U	8.62	6.94	5.46	0.51
	G2253A	1.11	0.00	0.00	0.00
	U2255C	0.87	0.26	0.22	0.15
Amplicón 3	U3402C	0.90	7.06	7.99	9.45
	C3545U	1.29	94.26	96.82	99.44
	A3603G	0.71	0.71	0.59	0.54

Tabla 35: Trayectoria seguida por las mutaciones representadas en la población ancestral en proporción mayor o igual al 0.5 % durante la evolución de la población G2_{P25.1}. Las celdas sombreadas en gris indican aumentos en la frecuencia de la mutación indicada (gris claro entre la población ancestral y el pase 50 y gris oscuro entre los pases 51 y 60), mientras que las sombreadas en rosa indican descensos de frecuencia (rosa claro entre la población ancestral y el pase 50 y rosa oscuro entre los pases 51 y 60). Las celdas correspondientes a mutaciones que no han alcanzado una representación del 5 % en ningún pase, o lo han hecho solo en uno, no se han resaltado.

En la línea G2_{P25.1}, durante los 50 pases realizados a temperaturas entre 38 °C y 42 °C hay 9 mutaciones de la población ancestral que aumentan claramente su representación (A1088G, U1295C, U1295G, G1312A, A2187C, C2201U, G2223A, U3402C, C3545U). Con las excepciones de A2187C (que aumenta solo en G2_{P25.1}) y C2228U (que aumenta solo en G1_{P25.2}), estas mutaciones son las mismas que aumentaban sus frecuencias durante la propagación a 40 °C en G1_{P25}. Todas ellas, excepto U1295C y C2228U (que decrecen) y C2201U (que se mantiene con ligeras fluctuaciones, continúan la tendencia ascendente durante los pases realizados a 43 °C. A pase número 60, la línea G2_{P25.1} presenta 9 mutaciones, de las ya identificadas en la población ancestral, representadas por encima del 5 %.

En conjunto, los resultados sugieren que los aumentos graduales de temperatura permiten que un número mayor de mutaciones de la población ancestral se mantengan a pase número 60 con frecuencias mayores del 5 % que en el caso de la línea C43_{P25.1}, evolucionada de forma constante a 43 °C a partir del mismo ancestro (4 mutaciones en este caso).

Trayectoria seguida por las mutaciones con representación mayor del 5 % que no estaban presentes en las poblaciones ancestrales

Las mutaciones que, sin estar presentes en la población ancestral con una frecuencia mayor del 0.5 %, alcanzan frecuencias mayores del 5 % durante la propagación de las líneas G1_{P25.2} y G2_{P25.1} aparecen en la Tabla 36.

		Frecuencia (%)				
		Qβ _{P25}	G1 _{P25.2} (P30)	G1 _{P25.2} (P31)	G1 _{P25.2} (P40)	G1 _{P25.2} (P60)
Amplícón 2	A2222C	0.18	0.07	0.10	7.61	0.15
	C2246U	0.07	5.65	7.23	16.89	57.50
	C2384U	0.20	5.60	7.23	15.35	57.23
	U2426C	0.07	10.32	11.11	2.45	0.30
	C2452U	0.09	0.12	0.29	2.50	5.81

		Frecuencia (%)			
		Qβ _{P25}	G2 _{P25.1} (P50)	G2 _{P25.1} (P51)	G2 _{P25.1} (P60)
Amplícón 1	A1149G	0.10	5.27	4.75	0.43
Amplícón 3	A3566G	0.21	4.82	6.51	7.49

Tabla 36: Mutaciones que no estaban presentes en la población ancestral en más del 0.5 % de las lecturas y que alcanzan una representación mayor o igual al 5 % durante la evolución de las poblaciones G1_{P25.2} y G2_{P25.1}.

En la línea G1_{p25.2}, todas estas mutaciones se encuentran localizadas en el amplicón 2. Dos de ellas (A2222C y U2426C) solo se mantienen por encima del 5 % de forma transitoria, mientras que las tres restantes aumentan gradualmente sus frecuencias, alcanzando sus máximos valores en el pase 60. En G2_{p25.1} solo aparecen dos nuevas mutaciones que alcancen una representación mayor del 5 % en algún pase. Son A1149G (lo alcanza de forma transitoria) y A3566G. Si se comparan estos resultados con los obtenidos para la población C43_{p25.1} se puede concluir que cuanto más gradual es el aumento de temperatura menos mutaciones se adquieren *de novo* y más se utiliza la diversidad preexistente en la población ancestral.

5.5.3 Variaciones de los haplotipos más representados a lo largo de la adaptación

De igual modo que para el caso de las líneas C43_{p2.3} y C43_{p25.1}, para analizar cómo se distribuyen las mutaciones, se determinaron las frecuencias de los cinco haplotipos más representados en los pases secuenciados mediante NGS de las líneas G1_{p25.2} y G2_{p25.1} y los comparamos con los cinco más representados en la correspondiente población ancestral (Tablas 37 y 38).

En la población C43_{p25.1} y en el amplicón 1, el haplotipo que más rápidamente se selecciona es el que lleva la mutación A1088G, inicialmente no acompañada de otras situadas en esta región y, más adelante, asociada a otras como G1312A o U1295G. Sin embargo, en ambas poblaciones graduales, el haplotipo que más aumenta su frecuencia en los pases previos a 43 °C es el que lleva únicamente la mutación G1312A, seguido de otros en los que esta mutación se asocia con otras como U1295G, U1295C o A1088G. Durante los pases a 43 °C, el haplotipo que solo lleva G1312A disminuye drásticamente su frecuencia en favor de los que llevan esa misma mutación en combinación con las mencionadas anteriormente.

En el amplicón 2, la diferencia más evidente entre los tres tipos de patrones de cambio es que en el caso de la línea evolucionada de forma constante el haplotipo con solo la mutación A2187C, que era el mayoritario en la población ancestral decae rápidamente su representación (a pase 10 es indetectable) en favor del que lleva la combinación A2222C + U2460C. Este último, más adelante pasa a ser sustituido por el que solo lleva A2222C. En las poblaciones graduales, sin embargo, el comportamiento de los haplotipos que contienen la mutación A2187C es claramente distinto. Durante los pases anteriores a 43 °C, en G1_{p25.2} el haplotipo con solo A2187C pasa de estar representado al 55 % a estarlo al 17 %, observándose que otros que llevan esta misma mutación acompañada de otras (como C2201U) aumentan sus frecuencias. En este tramo de la evolución el haplotipo más representado es el que contiene C2201U y no se observa el ascenso de los haplotipos con la mutación A2222C como se veía en la población constante, aunque sí lo hace durante los 10 primeros pases a 43 °C.

		G1 _{P25}							
		Q β _{P25}		G1 _{P25.2} (P30)		G1 _{P25.2} (P40)		G1 _{P25.2} (P60)	
Mutaciones		Posición	%	Posición	%	Posición	%	Posición	%
Amplicón 1	Ninguna	1	86.38	5	2.15	15	0.60	38	0.08
	G1312A	2	2.34	1	64.25	2	18.61	3	6.40
	U1295C	3	1.26	2	11.79	18	0.45	-	-
	U1328G	4	0.99	-	-	-	-	-	-
	A1088G	5	0.52	8	1.33	8	2.66	18	0.22
	U1295C,G1312A	-	-	3	4.40	5	2.92	9	0.99
	G1312A,A1324C	7	0.44	4	3.33	13	0.86	42	0.08
	A1088G,G1312A	-	-	6	1.92	1	35.00	1	46.66
	U1295G,G1312A	55	0.08	9	0.98	3	16.92	2	25.30
	U1295G	6	0.45	7	1.69	4	6.19	5	2.29
	A1088G,U1295G, G1312A	-	-	-	-	7	2.66	4	5.23
Amplicón 2	A2187C	1	54.97	2	17.04	1	28.11	2	18.64
	Ninguna	2	14.79	10	1.06	13	0.71	22	0.3
	C2228U	3	8.10	4	8.77	7	4.69	10	1.68
	A2187C,C2201U	4	3.41	5	7.48	4	7.29	4	4.63
	C2201U	5	2.42	1	34.46	2	23.86	3	6.91
	A2187C,C2201U, U2426C	-	-	3	9.51	-	-	-	-
	A2187C,C2246U, C2384U	-	-	6	4.81	3	7.68	1	39.58
	A2222C	59	0.08	-	-	5	7.21	81	0.07
	A2187C,C2201U, C2246U,C2384U	-	-	-	-	-	-	5	4.06
Amplicón 3	Ninguna	1	89.93	1	46.12	2	22.44	2	6.64
	C3545U	2	1.29	2	41.82	1	62.62	1	76.06
	U3402C	3	0.90	3	5.37	7	0.94	51	0.08
	A3603G	4	0.71	5	0.66	12	0.20	58	0.07
	C3517U	5	0.32	13	0.13	33	0.08	-	-
	U3402C,C3545U	-	-	4	0.73	6	1.18	3	5.37
	A3335G,C3545U	-	-	-	-	3	4.49	4	2.04
	C3338U,C3545U	-	-	6	0.39	4	1.65	-	-
	U3447C	-	-	-	-	5	1.44	-	-
	G3466A,C3545U	-	-	-	-	17	0.14	5	1.70

Tabla 37: Secuencias de los 5 haplotipos más representados a lo largo de la adaptación de la línea G1_{P25.2}. En cada columna aparece el número de orden (posición) de cada haplotipo y su frecuencia (%). Las mutaciones indicadas para cada haplotipo indican cambios respecto al virus Q β _{WT}. Los cinco haplotipos más representados en cada pase aparecen sombreados. La posición de estos haplotipos también fue rastreada en el resto de los pases y, en el caso de haber sido localizados entre los contenidos en el umbral del 0.05 %, aparece indicada. Los casos en los que el haplotipo no ha sido encontrado en otras poblaciones se denotan con -.

		G2 _{P25}					
		Q β _{P25}		G2 _{P25.1} (P50)		G2 _{P25.1} (P60)	
	Mutaciones	Posición	%	Posición	%	Posición	%
Amplicón 1	Ninguna	1	86.38	19	0.31	52	0.07
	G1312A	2	2.34	1	35.67	4	2.81
	U1295C	3	1.26	7	2.22	25	0.14
	U1328G	4	0.99	-	-	-	-
	A1088G	5	0.52	-	-	-	-
	U1295G,G1312A	55	0.08	2	22.51	1	70.00
	U1295C,G1312A	-	-	3	15.73	2	8.80
	A1088G,G1312A	-	-	4	5.43	3	6.24
	A1149G,U1295C, G1312A	-	-	5	3.31	13	0.21
	A1088G,U1295C, G1312A	-	-	11	0.53	5	2.03
	U1295G	6	0.45	29	0.15	12	0.27
Amplicón 2	A2187C	1	54.97	1	66.83	1	68.07
	Ninguna	2	14.79	12	0.66	21	0.24
	C2228U	3	8.10	2	6.15	13	0.38
	A2187C,C2201U	4	3.41	3	6.06	2	7.98
	C2201U	5	2.42	-	-	-	-
	A2187C,G2223A	11	0.74	4	3.43	3	7.25
	A2187C,G2217A	10	0.86	5	1.69	11	0.42
	A2187C,C2309U	-	-	24	0.20	4	2.27
	A2187C,C2452U	49	0.09	14	0.45	5	1.54
Amplicón 3	Ninguna	1	89.93	3	5.52	4	0.56
	C3545U	2	1.29	1	74.48	1	75.62
	U3402C	3	0.90	26	0.15	-	-
	A3603G	4	0.71	56	0.07	-	-
	C3517U	5	0.32	-	-	-	-
	U3402C,C3545U	-	-	2	6.91	2	9.03
	C3545U,A3566G	-	-	4	4.48	3	7.22
	C3545U,A3603G	-	-	5	0.64	5	0.54

Tabla 38: Secuencias de los 5 haplotipos más representados a lo largo de la adaptación de la línea G2_{P25.1}. En cada columna aparece el número de orden (posición) de cada haplotipo y su frecuencia (%). Las mutaciones indicadas para cada haplotipo indican cambios respecto al virus Q β _{WT}. Los cinco haplotipos más representados en cada pase aparecen sombreados. La posición de estos haplotipos también fue rastreada en el resto de los pases y, en el caso de haber sido localizados entre los contenidos en el umbral del 0.05 %, aparece indicada. Los casos en los que el haplotipo no ha sido encontrado en otras poblaciones se denotan con -.

Durante el tramo de evolución a 43 °C, los haplotipos en los que A2187C está combinada con otras mutaciones, en general aumentan sus frecuencias, salvo la combinación

A2187C + C2201U que las disminuye ligeramente. El haplotipo mayoritario al fin de la evolución a 40 °C, C2201U, también disminuye sus frecuencias durante la evolución a 43 °C. En G2_{P25.1}, el haplotipo que contiene únicamente la mutación A2187C continúa siendo el mayoritario durante toda la evolución a temperaturas mayores de 37 °C, aumentando desde el 55 % hasta el 68 %.

Por último, en el amplicón 3, la evolución constante a 43 °C no parece favorecer ninguna mutación en el caso analizado. Sin embargo, en las dos pautas de evolución gradual, el haplotipo conteniendo C3545U parece ser el más favorecido, seguido en bastante menor proporción del que lleva esa misma mutación combinada con U3402C.

5.5.4 Determinación de la diversidad genética en las poblaciones G1_{P25.2} y G2_{P25.1} mediante la estimación de índices de diversidad

Índices de diversidad basados en incidencia

Los valores de los índices de diversidad basados en la incidencia de mutaciones (*F_{mín}*) o de haplotipos diferentes (*dH_p*) se muestran en la Tabla 39. En todos los casos también se incluyen los valores obtenidos para la población ancestral Q β _{P25}, de modo que se pueda seguir su variación desde el origen. Puesto que los pases secuenciados para las poblaciones evolucionadas con distintas pautas de cambio de temperatura no fueron los mismos, se decidió comparar únicamente los valores obtenidos en el pase número 60 (Fig. 46).

		N	Hpl	dHpl($\cdot 10^{-4}$)	nM	F _{mín} ($\cdot 10^{-7}$)
Amplicón 1	Q β _{P25}	235849	84	3.56	81	12.63
	G1 _{P25.2} (P30)	237588	76	3.20	67	10.37
	G1 _{P25.2} (P31)	207149	73	3.52	65	11.54
	G1 _{P25.2} (P40)	244983	45	1.84	25	3.75
	G1 _{P25.2} (P60)	201042	60	2.98	41	7.50
Amplicón 2	Q β _{P25}	279452	71	2.54	56	6.09
	G1 _{P25.2} (P30)	330443	64	1.94	36	3.31
	G1 _{P25.2} (P31)	261548	75	2.87	41	4.76
	G1 _{P25.2} (P40)	310689	55	1.77	29	2.84
	G1 _{P25.2} (P60)	289525	85	2.94	48	5.04
Amplicón 3	Q β _{P25}	351353	75	2.13	74	7.66
	G1 _{P25.2} (P30)	212969	52	2.44	34	5.80
	G1 _{P25.2} (P31)	275050	54	1.96	35	4.63
	G1 _{P25.2} (P40)	238890	49	2.05	37	5.63
	G1 _{P25.2} (P60)	242864	70	2.88	65	9.73

		N	Hpl	dHpl($\cdot 10^{-4}$)	nM	Fmín($\cdot 10^{-7}$)
Amplicón 1	Q β _{P25}	235849	84	3.56	81	12.63
	G2 _{P25.1} (P50)	237433	62	2.61	36	5.57
	G2 _{P25.1} (P51)	263451	57	2.16	33	4.60
	G2 _{P25.1} (P60)	276971	70	2.53	57	7.57
Amplicón 2	Q β _{P25}	279452	71	2.54	56	6.09
	G2 _{P25.1} (P50)	282673	72	2.55	65	6.99
	G2 _{P25.1} (P51)	242113	77	3.18	71	8.91
	G2 _{P25.1} (P60)	242798	71	2.92	65	8.14
Amplicón 3	Q β _{P25}	351353	75	2.13	74	7.66
	G2 _{P25.1} (P50)	228133	70	3.07	65	10.36
	G2 _{P25.1} (P51)	331820	64	1.93	60	6.57
	G2 _{P25.1} (P60)	289698	65	2.24	61	7.66

Tabla 39: Valores de los índices de diversidad basados en incidencia para las líneas evolutivas G1_{P25.2}, G2_{P25.1} y su correspondiente ancestro Q β _{P25}. La definición exacta de los parámetros mostrados en la tabla se encuentra en el apartado 4.2.7. de Materiales y Métodos.

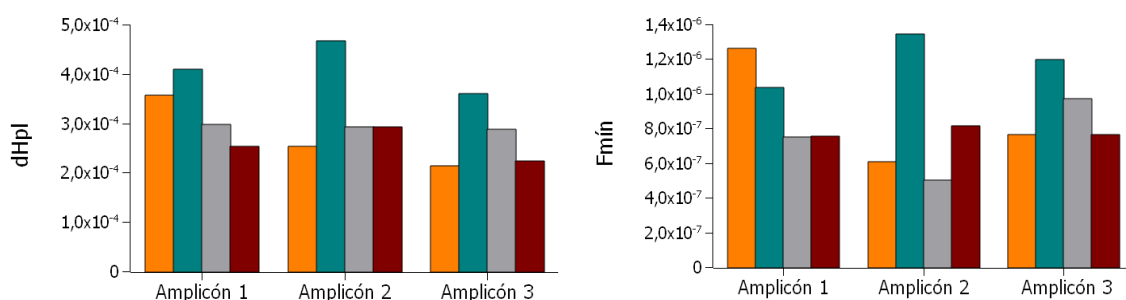


Figura 46: Densidad de haplotipos y frecuencia de mutación mínima para el ancestro Q β _{P25} (●) y las líneas C43_{P25.1} (●), G1_{P25.2} (●) y G2_{P25.1} (●) a pase 60.

Se observa que, mientras que en las tres regiones analizadas de la línea C43_{P25.1} hay incrementos en la densidad de haplotipos respecto a la población ancestral, en las poblaciones que han evolucionado con una pauta de cambio gradual se alcanzan valores menores, que en la mayoría de los casos son inferiores o solo ligeramente superiores a los de la población ancestral.

Los valores de *Fmín* muestran un comportamiento bastante similar al de la distribución de haplotipos, siendo en todas las regiones más altos para la población evolucionada de forma constante a 43 °C que para las que han seguido un patrón de cambio gradual. También se observa que estas últimas casi en ningún caso llegan a superar los valores de la población ancestral.

Índices de diversidad basados en abundancia

En la Tabla 40 y la Fig. 47 se muestran los valores de $F_{\text{máx}}$, π , H_s , H_sH y H_sN para las mismas poblaciones, regiones genómicas y número de pase del apartado anterior. Con pocas excepciones, la línea $G1_{P25.2}$ es la que muestra valores mayores para todos los índices considerados, siendo esto mucho más notable en el caso del amplicón 2. La línea $G2_{P25.1}$, en general, presenta menos diferencias con respecto a la línea evolucionada de forma constante.

		N	Hs	HsN($\cdot 10^{-2}$)	HsH	π ($\cdot 10^{-3}$)	$F_{\text{máx}}$ ($\cdot 10^{-4}$)
Amplicón 1	$Q\beta_{P25}$	235849	0.91	7.36	0.21	1.04	5.23
	$G1_{P25.2}$ (P30)	237588	1.65	13.32	0.38	3.34	43.06
	$G1_{P25.2}$ (P31)	207149	1.67	13.63	0.39	3.33	43.8
	$G1_{P25.2}$ (P40)	244983	2.18	17.54	0.57	5.23	67.63
	$G1_{P25.2}$ (P60)	201042	1.86	15.22	0.45	4.52	75.75
Amplicón 2	$Q\beta_{P25}$	279452	1.91	15.24	0.45	3.15	30.53
	$G1_{P25.2}$ (P30)	330443	2.31	18.21	0.56	5.74	46.47
	$G1_{P25.2}$ (P31)	261548	2.44	19.57	0.56	6.02	48.38
	$G1_{P25.2}$ (P40)	310689	2.37	18.78	0.59	6.68	48.76
	$G1_{P25.2}$ (P60)	289525	2.42	19.21	0.54	6.29	75.57
Amplicón 3	$Q\beta_{P25}$	351353	0.72	5.61	0.17	0.73	3.66
	$G1_{P25.2}$ (P30)	212969	1.30	10.60	0.33	2.65	20.79
	$G1_{P25.2}$ (P31)	275050	1.33	10.62	0.33	2.71	22.28
	$G1_{P25.2}$ (P40)	238890	1.33	10.77	0.34	2.47	32.24
	$G1_{P25.2}$ (P60)	242864	1.23	9.91	0.29	1.71	40.09

		N	Hs	HsN($\cdot 10^{-2}$)	HsH	π ($\cdot 10^{-3}$)	$F_{\text{máx}}$ ($\cdot 10^{-4}$)
Amplicón 1	$Q\beta_{P25}$	235849	0.91	7.36	0.21	1.04	5.23
	$G2_{P25.1}$ (P50)	237433	2.14	17.30	0.52	4.18	62.32
	$G2_{P25.1}$ (P51)	263451	2.13	17.09	0.53	4.30	65.89
	$G2_{P25.1}$ (P60)	276971	1.42	11.38	0.34	2.70	75.82
Amplicón 2	$Q\beta_{P25}$	279452	1.91	15.24	0.45	3.15	30.53
	$G2_{P25.1}$ (P50)	282673	1.68	13.37	0.39	2.64	39.13
	$G2_{P25.1}$ (P51)	242113	1.82	14.69	0.42	2.55	39.66
	$G2_{P25.1}$ (P60)	242798	1.57	12.64	0.37	2.09	40.83
Amplicón 3	$Q\beta_{P25}$	351353	0.72	5.61	0.17	0.73	3.66
	$G2_{P25.1}$ (P50)	228133	1.26	10.21	0.30	1.82	41.67
	$G2_{P25.1}$ (P51)	331820	1.19	9.39	0.29	1.75	43.23
	$G2_{P25.1}$ (P60)	289698	1.14	9.06	0.27	1.71	45.03

Tabla 40: Valores de los índices de diversidad basados en abundancia para las líneas evolutivas $G1_{P25.2}$, $G2_{P25.1}$ y su correspondiente ancestro $Q\beta_{P25}$. La definición exacta de los parámetros mostrados en la tabla se encuentra en el apartado 4.2.7. de Materiales y Métodos.

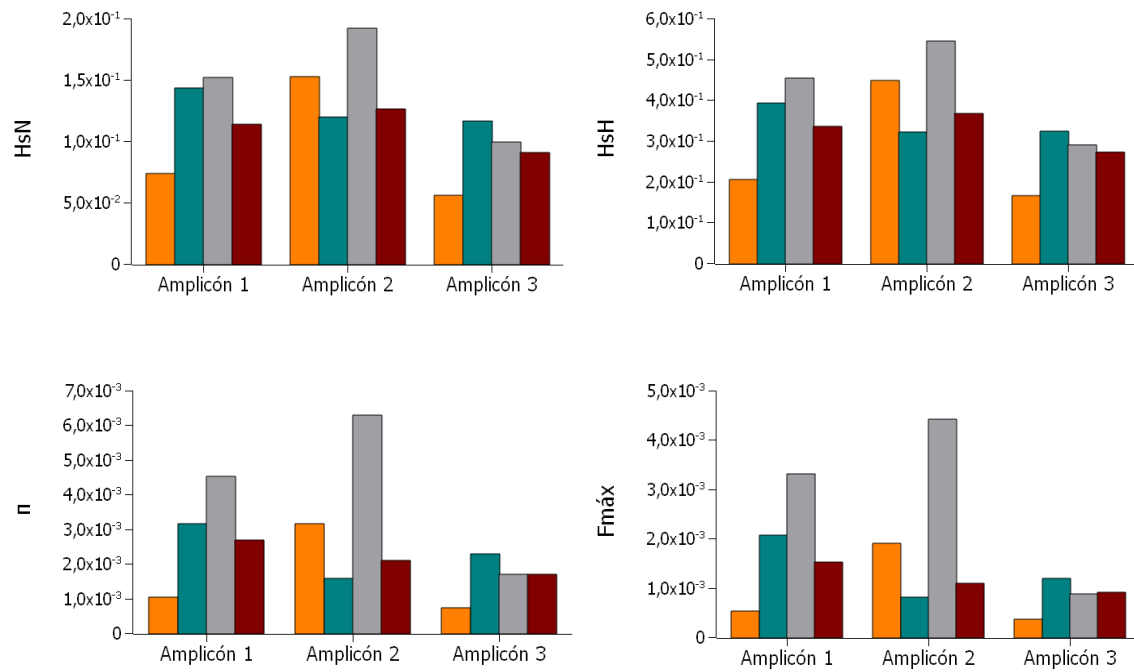


Figura 47: Valores de los índices de diversidad basados en abundancia para el ancestro $Q\beta_{P25}$ (●) y las líneas $C43_{P25.1}$ (●), $G1_{P25.2}$ (●) y $G2_{P25.1}$ (●) a pase 60.

5.6 CAPÍTULO 6

Valor selectivo de las mutaciones seleccionadas durante la adaptación a temperatura elevada

5.6.1 Detección de mutaciones en la secuencia consenso en función de la temperatura de evolución

Las poblaciones sometidas a los incrementos progresivos de temperatura ($G1_{P25}$ y $G2_{P25}$) evolucionaron a través de una serie de diferentes ambientes, cada uno de ellos caracterizado por un valor de temperatura. El análisis de las mutaciones seleccionadas en cada uno de estos entornos puede dar información sobre la temperatura a la cual cada mutación ofrece mayores ventajas selectivas. Podría ocurrir que algunas de las mutaciones seleccionadas a temperaturas más moderadas tuvieran efectos negativos a temperaturas más altas, en cuyo caso no podrían mantenerse a no ser que surgieran otras con efectos compensatorios. También podría haber mutaciones que proporcionaran un contexto genómico en el que la probabilidad de fijación de otras adicionales se viera modificada, tanto en forma positiva como negativa.

Por el contrario, las líneas expuestas directamente a 43 °C ($C43_{P2}$ y $C43_{P25}$) se ven forzadas a seleccionar las mutaciones que ofrecen ventajas selectivas a esta temperatura. En el caso de que los efectos beneficiosos a 43 °C de alguna mutación fueran dependientes de la presencia de otras que solo se pueden seleccionar a temperaturas más bajas, esas mutaciones no podrían aparecer en las líneas propagadas de forma constante a 43 °C.

Cuando se consideran las mutaciones detectadas en el rango entre 38 °C y 40 °C en los conjuntos de líneas evolutivas $G1_{P25}$ y $G2_{P25}$, y que están representadas al menos en la mitad de las líneas, se encuentran las mutaciones siguientes: U1295C/G, G1312A, G1371A, C1806U, A1930G, A2187C, C2201U, U2776C, C3545U Y G3945A (Tabla 41). Algunas de estas mutaciones también estaban presentes en al menos tres líneas de los dos patrones de propagación constante a 43 °C, $C43_{P2}$ y $C43_{P25}$, (U1295C/G, G1312A, G1371A, A2187C, U2776C y G3945A), lo que sugiere que probablemente son beneficiosas en todo el rango de temperaturas entre 38 °C y 43 °C. En contraste con estas observaciones, C1806U, A1930G y C2201U no fueron detectadas en ninguna de las líneas $C43_{P25}$ ni $C43_{P2}$, lo que indica que probablemente producen sus mayores aumentos en *fitness* a temperaturas entre 38 °C y 40 °C, pudiendo incluso tener efectos negativos a 43 °C, algo que parece ocurrir en el caso de las sustituciones A1930G y C2201U que mostraron una clara tendencia a desaparecer a pases posteriores en los que la temperatura se elevó por encima de 40 °C. C3545U no estuvo totalmente ausente en las líneas evolucionadas a 43 °C, ya que apareció a pases tardíos como polimorfismo en la línea $C43_{P2.1}$. La mayoría de estas mutaciones, excepto G3945A en una línea, no están entre las primeras fijadas en las líneas $C43_{P25}$. Sí hay varias entre las primera fijadas en las líneas $C43_{P2}$.

Mutación	Línea evolucionada			
	G1 _{P25}	G2 _{P25}	C43 _{P2}	C43 _{P25}
U1295C/G	2	1	1	3
G1312A	3	3	3	3
G1371A	1	2	3	3
C1806U	3	3	0	0
A1930G	3	3	0	0
A2187C	2	3	3	0
C2201U	3	3	0	0
U2776C	3	3	2	2
C3545U	3	3	1	0
G3945A	2	1	2	3

Tabla 41: Número de veces que aparece representada cada mutación en los conjuntos de líneas evolutivas G1_{P25}, G2_{P25}, C43_{P2} y C43_{P25}. Las mutaciones mostradas corresponden a aquellas que aparecen por primera vez a temperaturas entre 38 °C y 40 °C en al menos 3 líneas del conjunto G1_{P25} + G2_{P25}. El sombreado muestra las mutaciones que también están presentes en al menos 3 líneas del conjunto C43_{P2} + C43_{P25}.

Cuando se consideran las mutaciones detectadas a temperatura mayor de 40 °C en los conjuntos de líneas evolutivas G1_{P25} y G2_{P25}, y que están representadas al menos en la mitad de las líneas, se encuentran las mutaciones A1088G, A1778G y G2223A (Tabla 42). De ellas, solo la mutación A1088G aparece mayoritariamente también en las líneas C43_{P2} y C43_{P25}.

Mutación	Línea evolucionada			
	G1 _{P25}	G2 _{P25}	C43 _{P2}	C43 _{P25}
A1088G	2	1	3	3
A1778G	2	2	0	0
G2223A	2	1	0	1

Tabla 42: Número de veces que aparece representada cada mutación en los conjuntos de líneas evolutivas G1_{P25}, G2_{P25}, C43_{P2} y C43_{P25}. Las mutaciones mostradas corresponden a aquellas que aparecen por primera vez a temperaturas superiores a 40 °C en al menos 3 líneas del conjunto G1_{P25} + G2_{P25}. El sombreado muestra la única mutación que también está presente en al menos 3 líneas del conjunto C43_{P2} + C43_{P25}.

La aparición por primera vez de las mutaciones A1088G, A1778G y G2223A a temperaturas superiores a 40 °C en las líneas evolutivas G1_{P25} y G2_{P25} sugiere que no tienen efectos ventajosos a temperaturas inferiores a 40 °C. La mutación A1088G también se seleccionó en todas las líneas evolutivas propagadas de forma constante a 43 °C. Sin embargo, A1778G no fue detectada en ninguna de estas líneas y G2223A solo apareció de forma transitoria en C43_{P25}.1, indicando que estas dos mutaciones podrían requerir la presencia de otras para poder permanecer a 43 °C.

Entre las mutaciones que aparecen en al menos 3 líneas de los conjuntos C43_{P2} y C43_{P25} hay algunas (U2006G, U2016C, A2222C, U3311G y C3903U) que no aparecen en ninguna de las líneas evolucionadas de forma gradual (Tabla 43). Esto podría ser debido a que el contexto genómico generado durante la evolución a temperaturas más bajas no favorece su selección. Otra posibilidad es que el número de pases experimentado por estas poblaciones a 43 °C no haya sido suficiente para que estas mutaciones puedan llegar a ser detectables en la secuencia consenso.

Mutación	Línea evolucionada			
	C43 _{P2}	C43 _{P25}	G1 _{P25}	G2 _{P25}
U2006G	2	1	0	0
U2016C	1	2	0	0
A2222C	1	3	0	0
U3311G	1	3	0	0
C3903U	1	3	0	0

Tabla 43: Número de veces que aparece representada cada mutación en las líneas evolutivas C43_{P2} y C43_{P25}, G1_{P25} y G2_{P25}. Las mutaciones mostradas corresponden a aquellas que aparecen en al menos 3 líneas de los conjuntos C43_{P2} y C43_{P25} y que no aparecen en ninguna de las líneas evolucionadas de forma gradual.

En conjunto, los resultados sugieren que hay mutaciones cuyos efectos beneficiosos solo ocurren a temperaturas superiores a 40 °C (A1088G, A1778G, U2006G, U2016C, A2222C, G2223A, U3311G y C3903U). Hay otras que parecen seleccionarse preferentemente a temperaturas inferiores a 40 °C (C1806U, A1930G, C2201U y C3545U). De ellas, algunas son seleccionadas en contra a 43 °C (A1930G y C2201U), lo que parece sugerir que tienen costes en *fitness* a esta temperatura, mientras que otras pueden permanecer a 43 °C (C1806U y C3545U). Por último parece que hay otras mutaciones con efectos beneficiosos a todas las temperaturas por encima de 37 °C (U1295C/G, G1312A, G1371A, A2187C, U2776C y G3945A).

5.6.2 Efectos en el *fitness* de las mutaciones seleccionadas a temperatura elevada cuando están presentes en el contexto genético del virus salvaje

Con objeto de determinar los efectos en el *fitness* a distintas temperaturas de las mutaciones seleccionadas en los experimentos de adaptación de Q β a temperatura elevada, se procedió a preparar mutantes únicos que contuvieran algunas de esas mutaciones. Las mutaciones seleccionadas fueron A1088G (que solo se selecciona a 43 °C), C1806U y C3545U (que parecen seleccionarse preferentemente a temperaturas moderadas y que pueden mantenerse a temperaturas elevadas), A1930G (que parece tener efectos beneficiosos a temperaturas moderadas y negativos a temperaturas de 40 °C o superiores), y, por último, G1312A y U2776C que parecen tener efectos beneficiosos a todas las temperaturas en el rango 38 - 43 °C.

La tasa de crecimiento de todos estos mutantes se determinó en un ensayo de *fitness* realizado a 37, 40 y 43 °C (Fig. 48).

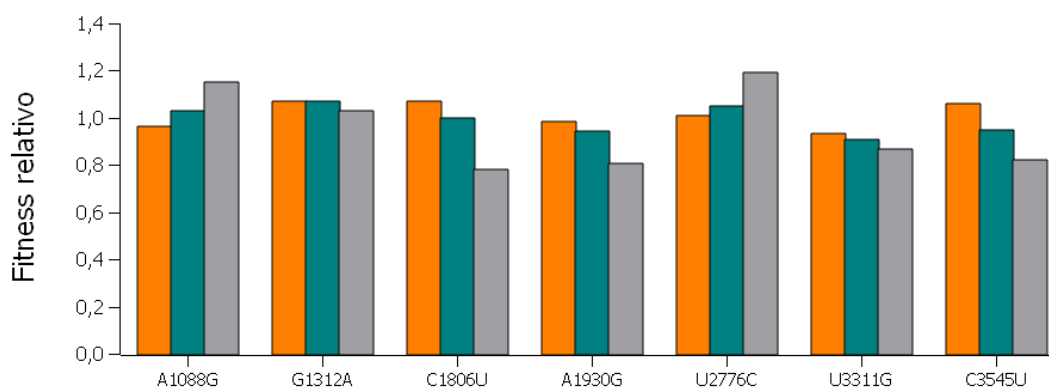


Figura 48: Valores de *fitness* a diferentes temperaturas de los mutantes conteniendo las mutaciones indicadas; 37 °C (●), 40 °C (●) y 43 °C (●). El virus de referencia correspondió a un clon biológico obtenido tras la expresión del plásmido pBRT7Q β .

En el caso del mutante A1088G se confirma que los mayores efectos beneficiosos de la mutación se producen a 43 °C, aunque son más leves de lo que podría esperarse dada la total coincidencia de esta mutación a pases tempranos en todas las líneas evolutivas que fueron expuestas directamente a 43 °C. En el caso de las mutaciones C1806U, A1930G y C3545U, se observa que son mutaciones perjudiciales a 43 °C y con efectos menos deletéreos (o incluso débilmente beneficiosos como en el caso de las mutaciones C1806U y C3545U a 37 °C) a otras temperaturas. El hecho de que tanto C1806U como C3545U puedan mantenerse a 43 °C sugiere que existen otras mutaciones que compensan sus efectos perjudiciales a esta temperatura. También se confirma que las mutaciones G1312A y U2776C son beneficiosas a todas las temperaturas ensayadas, aunque solo U2776C tiene efectos claros en el *fitness* a 43 °C. El caso más discrepante es el de mutación U3311G, que está entre las mutaciones que se fijan en primer lugar en todas las líneas expuestas directamente a 43 °C con ancestro en Q β _{P25}, siendo por

tanto candidata a tener efectos beneficiosos a esta temperatura, algo que no se observa en los valores obtenidos para su tasa de crecimiento. Este resultado sugiere que sus efectos beneficiosos probablemente son debidos a su interacción con otras mutaciones. En cualquier caso, hay que señalar que no se observan grandes diferencias en los valores de *fitness* de ninguno de los mutantes respecto a los observados para el virus salvaje, lo que sugiere que las ventajas adaptativas de todas estas mutaciones se deben a su combinación con otras en el mismo genoma.

5.7 CAPÍTULO 7

Adaptación del bacteriófago Q β a temperaturas alternantes

Hasta ahora se ha estudiado cómo influyen la diversidad genética preexistente en la población inicial y la pauta de cambio de temperatura en la adaptación del bacteriófago Q β a replicar a 43 °C. Sin embargo, lo normal en la naturaleza es que los valores de la mayoría de las presiones selectivas (la temperatura entre ellas) experimenten fluctuaciones de diversa intensidad, por encima y por debajo de un valor óptimo. Para estudiar cómo podrían afectar ese tipo de variaciones a la evolución de Q β , se diseñó un experimento en el que la población ancestral Q β_{P2} se propagó por triplicado durante 60 pases seriados que alternaron las temperaturas de 43 °C y 30 °C, originándose así las líneas evolutivas A $_{P2.1}$, A $_{P2.2}$ y A $_{P2.3}$. Por problemas técnicos, una de las líneas evolutivas tuvo que ser descartada, de modo que solo se presentarán los resultados obtenidos para A $_{P2.1}$ y A $_{P2.2}$. En paralelo, la población Q β_{P2} también se propagó durante 60 pases a la temperatura constante de 30 °C, dando lugar a las líneas evolutivas C30 $_{P2.1}$, C30 $_{P2.2}$ y C30 $_{P2.3}$. Como control de la replicación de forma constante a 43 °C se utilizaron las líneas C43 $_{P2}$ descritas en los capítulos 2 y 3.

5.7.1 *Fitness* a 30 °C y 43 °C de las poblaciones evolucionadas de forma constante y alternante a esas temperaturas

Con objeto de analizar el efecto de la alternancia de la replicación a 30 °C y 43 °C en la evolución del virus, en primer lugar se realizó un ensayo de *fitness* en el que se determinó la tasa de crecimiento de las líneas evolutivas obtenidas a pase 60. La temperatura del ensayo fue 30 °C para las líneas C30 $_{P2}$ y A $_{P2}$ y de 43 °C para las líneas C43 $_{P2}$ y A $_{P2}$. Los resultados obtenidos, normalizados por el valor de la población ancestral Q β_{P2} , se muestran en la Fig. 49.

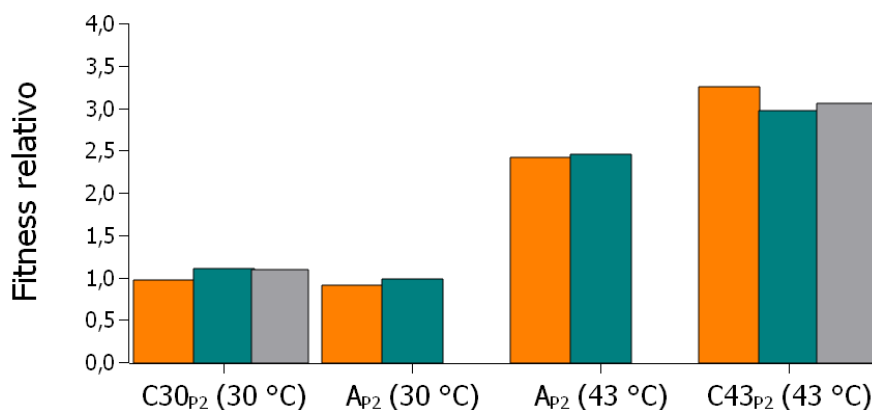


Figura 49: Valores de *fitness* de las líneas evolutivas C30 $_{P2}$, A $_{P2}$ y C43 $_{P2}$ en el pase número 60. Los ensayos de *fitness* se realizaron a la temperatura indicada en cada caso. (●) réplica 1, (●) réplica 2 y (●) réplica 3. El virus de referencia utilizado fue la población ancestral Q β_{P2} que es el origen de todas las líneas evolutivas representadas.

Se observa que, mientras que los valores de *fitness* a 30 °C apenas experimentan cambios en ningún caso, los valores de *fitness* a 43 °C experimentan aumentos relevantes, que son más evidentes en el caso de las líneas C43_{P2}.

A continuación, se determinaron los valores de *fitness* a 43 °C cada diez pases para las líneas A_{P2}. De este modo se pudo comparar su dinámica de variación de *fitness* con las de las líneas propagadas de forma constante a 43 °C.

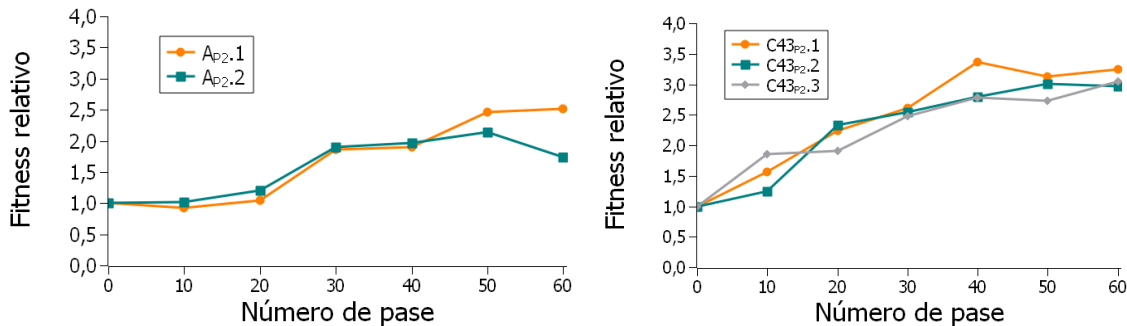


Figura 50: Dinámica de variación del *fitness* a 43 °C en las líneas evolutivas A_{P2} y C43_{P2}. El virus de referencia utilizado fue la población ancestral Qβ_{P2} que es el origen de todas las líneas evolutivas representadas.

Se observa que las dos líneas alternantes apenas experimentan variaciones de *fitness* hasta el pase número 20. A partir de ese punto, y hasta el pase número 30, es cuando se producen las mayores ganancias de *fitness*. Esta dinámica es muy diferente de la observada en las poblaciones evolucionadas de forma constante a 43 °C, en las que se observaron aumentos lineales de *fitness* durante los primeros 40 pases.

A continuación, se estimaron los valores de *fitness* alcanzados a pase 60, las diferencias de *fitness* alcanzadas durante los primeros 10 pases y el número de pases necesarios para duplicar el valor de *fitness* de la población ancestral. Estos valores se compararon con los de la población evolucionada de forma constante a 43 °C.

Línea evolutiva	Número de pase (fitness relativo = 2)	Fitness relativo a pase 60	Diferencia de fitness (primeros 10 pases a 43 °C)
C43 _{P2} .1	16.3	3.3	0.6
C43 _{P2} .2	16.9	3.0	0.3
C43 _{P2} .3	21.5	3.1	0.9
C43 _{P2} (Media ± SD)	18.2 ± 2.8	3.1 ± 0.1	0.6 ± 0.3
A _{P2} .1	41.8	2.5	0.0
A _{P2} .2	42.2	1.7	0.2
A _{P2} (Media ± SD)	42.0 ± 0.3	2.1 ± 0.5	0.1 ± 0.1

Tabla 44: Parámetros cuantitativos que describen la dinámica de la variación del *fitness* a 43 °C en las líneas evolutivas C43_{P2} y A_{P2}. Media y SD hacen referencia a los valores medios y su correspondiente desviación estándar. Debido a la alternancia de temperaturas en cada pase, en el caso de las poblaciones A_{P2} los diez pases a 43 °C no se alcanzan hasta el pase número 20.

Aunque, en el caso de las poblaciones alternantes, la interpolación en las gráficas de variación de *fitness* (Fig. 50) indica que el número de pases requeridos para duplicar el valor de *fitness* inicial es muy elevado, en realidad, a pase 30 ya casi se alcanza un valor de *fitness* relativo de 2. Las ganancias de *fitness* durante los 10 primeros pases a 43 °C también son más bajas para las líneas alternantes que para las evolucionadas de forma constante a 43 °C. Por último el valor de *fitness* alcanzado a pase número 60 fue significativamente menor. Sin embargo, a la hora de interpretar este resultado hay que tener en cuenta que en las poblaciones A_{p2} la mitad de los pases se realizaron a 30 °C.

5.7.2 Mutaciones adquiridas durante la evolución a temperaturas constantes y alternantes

Las mutaciones presentes en las secuencias consenso de las líneas evolutivas A_{p2}, C30_{p2} y C43_{p2} a pase número 60 aparecen en la Tabla 45. La comparación de las mutaciones fijadas o polimórficas en estos conjuntos de líneas permitió analizar las similitudes y las diferencias entre ellas.

	U458C	U507C	U612C	C714U	A815G	A1088G	G1122A	C1228U	U1295C/G	G1312A	G1371A	C1649U
C30 _{p2}												
A _{p2}	X	X	X	X	X	X	X	X	X	X	X	X
C43 _{p2}												

	C1803U	C1806U	C1807U	C1810U	G1820A	U2006G	C2011A	U2016C	G2117A	A2187C	C2210U	G2211A	A2222C
C30 _{p2}													
A _{p2}	X	X	X	X	X	X	X	X	X	X	X	X	X
C43 _{p2}													

	U2426C	C2452U	G2741A	U2776C	U3311G	C3344U	U3402C	C3413U	U3447C	C3545U	C3884U	C3903U	G3945A
C30 _{p2}													
A _{p2}	X	X	X	X	X	X	X	X	X	X	X	X	X
C43 _{p2}													

wt	mut<wt	mut=wt	mut>wt	mut
----	--------	--------	--------	-----

Tabla 45: Mutaciones y polimorfismos en las secuencias consenso de las líneas evolutivas C30_{p2}, A_{p2} y C43_{p2} en el pase número 60 de la adaptación. La primera fila muestra los diferentes cambios nucleotídicos observados. El estado de las mutaciones se muestra con un sombreado de diferente intensidad (ver parte inferior de la tabla) en los cuadrados situados debajo de cada mutación (de izquierda a derecha, réplicas 1, 2 y 3 respectivamente). La posición 1295 experimentó cambios U/C o U/G que, con frecuencia, coexistían en la misma población. En la figura no se ha hecho distinción entre ambos tipos de cambio. El recuadro correspondiente a la línea A_{p2}.3 (que se perdió) aparece con signo X.

El conjunto de todas las líneas evolutivas obtenidas a pase número 60 mostraron un total de 39 mutaciones diferentes que afectaron 38 posiciones nucleotídicas (Tabla 45). Al igual que se hizo en análisis previos, aunque la posición 1295 puede experimentar cambios U/C o U/G, dada la frecuente coexistencia de ambos cambios en la misma población, en este análisis no se distinguirá entre las dos situaciones.

Mutación ¹	Gen ²	Número de líneas con la mutación ³		
		C30 _{P2}	A _{P2}	C43 _{P2}
U458C (V132C)	A2	1/0		
U507C (S)	A2		1/0	
U612C (S)	A2		1/0	
C714U (S)	A2			0/1
A815G (N251S)	A2	0/1		
A1088G (D342G)	A2		2/0	2/1
G1122A (S)	A2	1/0	1/0	
C1228U (L389F)	A2			0/1
U1295(C/G) (F411S/C)	A2			1/0
G1312A (V417I)	A2		1/0	2/0
G1371A (G9S)	C		2/0	3/0
C1649U (S)	C		1/0	
C1803U (P153S)	A1	1/0		
C1806U (P154S)	A1	0/1		
C1807U (P154L)	A1			1/0
U1810C (I155T)	A1			1/0
G1820A (S)	A1			1/0
U2006G (S220R)	A1			0/2
C2011A (T222N)	A1	3/0		
U2016C (F224L)	A1		2/0	1/0
G2117A (S)	A1		1/0	
A2187C (S281R)	A1		1/0	0/3
C2201U (S)	A1	0/1		
G2211A (D289N)	A1	1/0		
A2222C (S)	A1		1/0	1/0
U2426C (S)	R		0/1	
C2452U (A33V)	R		1/0	1/0
G2741A (M129I)	R	1/0		
U2776C (V141A)	R		0/1	2/0
U3311G (I320M)	R		1/0	1/0
C3344U (S)	R		1/0	
U3402C (S350P)	R			1/0
C3413U (S)	R		1/0	
U3447C (C365R)	R		1/0	0/1
C3545U (S)	R	1/0		0/1
C3884U (S)	R		1/0	
C3903U (L517F)	R		1/1	0/1
G3945A (G531S)	R	0/1	2/0	2/0

Tabla 46: Mutaciones presentes en las líneas evolutivas C30_{P2}, A_{P2} y C43_{P2} en el pase número 60 de la adaptación.¹Todos los cambios se identificaron respecto a la secuencia del virus Qβ_{WT}. El cambio de aminoácido se muestra entre paréntesis. S indica que la mutación es sinónima. La posición 1295 puede mutar a C o G. ²Ubicación de la mutación en los diferentes genes del genoma de Qβ (C: cápsida, R: replicasa). ³El primer número de cada par indica el número de líneas en un conjunto dado en el que se fija la mutación. El segundo número se refiere al número de líneas en las que la mutación está presente como polimorfismo. Las celdas vacías indican la ausencia de la mutación en las líneas de ese conjunto en particular.

Del conjunto de 38 posiciones mutadas, 12 de ellas fueron comunes a las líneas A_{P2} y $C43_{P2}$ (A1088G, G1312A, G1371A, U2016C, A2187C, A2222C, C2452U, U2776C, U3311G, U3447C, C3903U y G3945A), mientras que solo 2 fueron comunes a las líneas A_{P2} y $C30_{P2}$ (G1122A y G3945A). La mutación C3545U apareció en los dos conjuntos de líneas $C30_{P2}$ y $C43_{P2}$, y solo G3945A fue común a todos ellos. La elevada coincidencia entre las mutaciones que aparecen en las líneas A_{P2} y $C43_{P2}$ podría indicar que muchas de las mutaciones que se seleccionan a 43 °C no tienen efectos negativos a 30 °C. Por el contrario, el hecho de que la mayoría de las mutaciones seleccionadas en las líneas $C30_{P2}$ no se detecten en las $C43_{P2}$ ni en las A_{P2} indicaría que no tienen efecto beneficioso a 43 °C o que incluso lo tienen deletéreo.

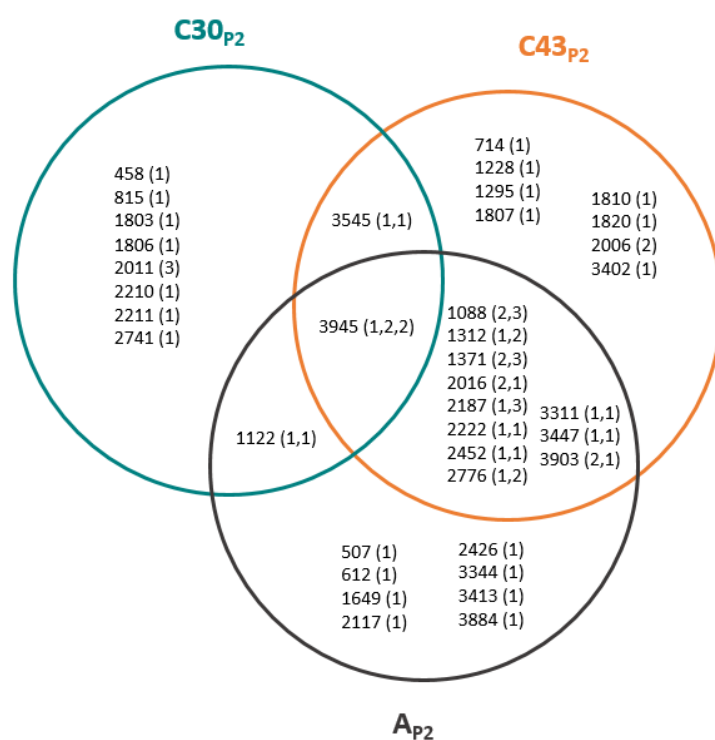


Figura 51: Diagrama de Venn mostrando las mutaciones coincidentes y exclusivas de las líneas evolutivas $C30_{P2}$, A_{P2} y $C43_{P2}$. La circunferencia correspondiente a cada conjunto de líneas contiene las posiciones nucleotídicas que aparecen mutadas a pase 60 en al menos una línea. En el caso de mutaciones coincidentes en los tres conjuntos, el primero de los números del paréntesis se refiere a las réplicas de $C30_{P2}$, el segundo a las de A_{P2} y el tercero a las de $C43_{P2}$. Para las mutaciones coincidentes en dos de los tres conjuntos, el orden de los número entre paréntesis es ($C30_{P2}$, A_{P2}), ($C30_{P2}$, $C43_{P2}$) y (A_{P2} , $C43_{P2}$).

Otra observación importante es que si se enumeran las mutaciones diferentes detectadas en las líneas evolucionadas de forma constante a 30 °C y 43 °C, se obtienen los valores de 11 y 21, indicando que la evolución a temperaturas mayores de la óptima va asociada a la aparición de un número mayor de mutaciones en la secuencia consenso que la evolución a temperaturas inferiores a la óptima. Las líneas alternantes también presentan 21 posiciones mutadas, pero en este caso, al disponer de solo dos líneas, los

números no son directamente comparables con los obtenidos para las líneas C30_{P2} o C43_{P2}.

Del total de mutaciones observadas a pase número 60, 8 fueron exclusivas de las líneas C43_{P2} (C714U, C1228U, U1295C/G, C1807U, C1810U, A1820G, U2006G y U3402C), 8 de las líneas A_{P2} (U507C, U612C, C1649U, G2117A, U2426C, C3344U, C3413U y C3884U) y 8 de las líneas C30_{P2} (U458C, A815G, C1803U, C1806U, C2011A, C2210U, G2211A y G2741A). Algunas de estas mutaciones exclusivas sí pudieron ser detectadas en otros experimentos de adaptación a temperatura elevada descritos en esta tesis, tanto cuando el aumento se produjo en un solo paso o de forma gradual. Este es el caso de las mutaciones C1228U, C1649U y U2006G que aparecían en algunas líneas de C43_{P25}, de la mutación U1295C/G que aparecía en C43_{P25}, G1_{P25} y G2_{P25} y de la mutación C1806U que aparecía en las líneas de los dos patrones de aumento de temperatura graduales, G1_{P25} y G2_{P25}.

En el conjunto total de mutaciones hay 14 sinónimas y 24 no sinónimas. De las sinónimas, 11 se localizan exclusivamente en una línea, mientras que las tres restantes aparecen en dos tratamientos, pero solo en una línea de cada uno de ellos. El número de mutaciones no sinónimas es de 24. De ellas también 11 son exclusivas de línea, mientras que el resto aparecen en varios tratamientos, a menudo representadas en más de una línea de cada uno de ellos. La conclusión es que un mayor número de mutaciones no sinónimas que sinónimas coinciden entre líneas y tratamientos. De hecho, de las 12 mutaciones que coinciden en las líneas A_{P2} y C43_{P2}, 11 son no sinónimas.

5.7.3 Estudio comparativo de la diversidad genética contenida en las poblaciones evolucionadas a temperaturas constantes o alternantes

Para comparar la estructura interna de las líneas evolutivas C43_{P2}, C30_{P2} y A_{P2}, se realizó un análisis mediante NGS de un representante de cada conjunto: C30_{P2.1}, A_{P2.2} y C43_{P2.3} (esta última, ya analizada en el Capítulo 3). Los procedimientos, regiones secuenciadas y umbral de representación de las secuencias utilizadas en el análisis fueron los mismos que para las poblaciones Q β _{P2} y Q β _{P25} (ver apartados 4.2.6., 4.2.7. y 5.1.1.). El número de lecturas totales, las obtenidas tras el procesado bioinformático y las correspondientes a las secuencias representadas en más del 0.05 % del total de lecturas se muestran en la Tabla 6.

Como una primera aproximación para analizar las diferencias entre las líneas evolutivas C30_{P2.1}, A_{P2.2} y C43_{P2.3}, se representaron las distribuciones de frecuencias de los haplotipos presentes en más del 0.05 % de las lecturas (Fig. 52). Para los tres amplicones estudiados se observa que el haplotipo más frecuente tiene los valores más bajos en la población C43_{P2.3}, mientras que los comprendidos entre aproximadamente las posiciones 2 y 25 alcanzan valores más altos en esta misma población que en las dos restantes. Las mayores diferencias se observan entre las líneas C30_{P2.1} y C43_{P2.3}.

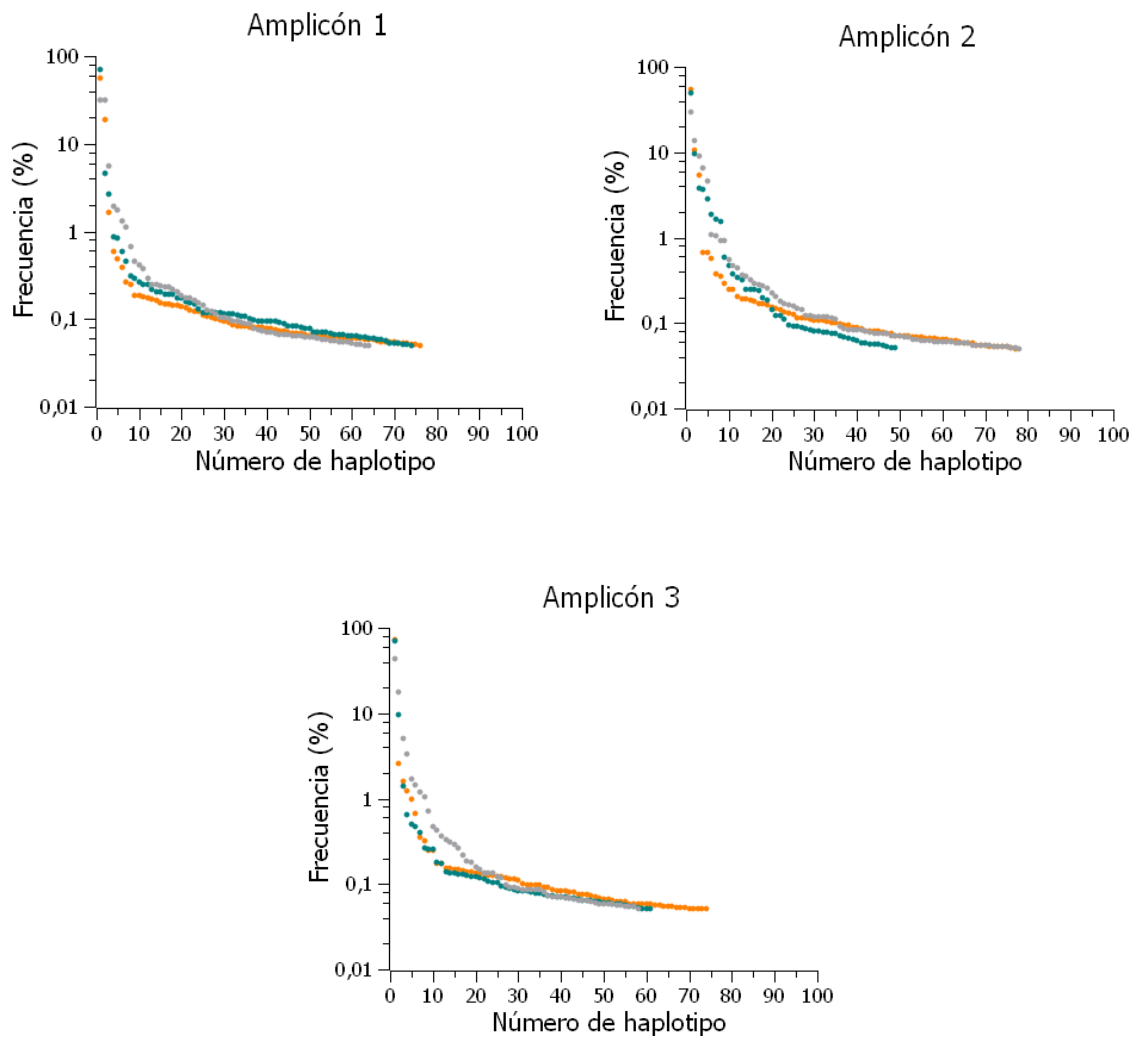


Figura 52: Frecuencias de los haplotipos representados en más del 0.05 % de las lecturas obtenidas a pase número 60 en cada uno de los amplicones analizados. Poblaciones C30_{p2.1} (●), Ap_{2.2} (●) y C43_{p2.3} (●). El número de cada haplotipo representa el número de orden en relación a su frecuencia.

La distribución de la frecuencia de haplotipos en función de su distancia de Hamming a la secuencia consenso se muestra en la Fig. 53. En todos los casos se observa que la frecuencia del haplotipo correspondiente a la secuencia consenso (distancia de Hamming 0) es menor para la línea C43_{p2.3} que para la línea C30_{p2.1}. Sin embargo, las frecuencias de los haplotipos a distancias de Hamming 1 o mayores son claramente mayores en el caso de la población C43_{p2.3}, lo cual sugiere que esta población es más diversa que la C30_{p2.1}. La línea Ap_{2.2} presenta menores diferencias con la línea C30_{p2.1}, sobre todo cuando se consideran los amplicones 2 y 3 y distancias de Hamming de 0 y 1.

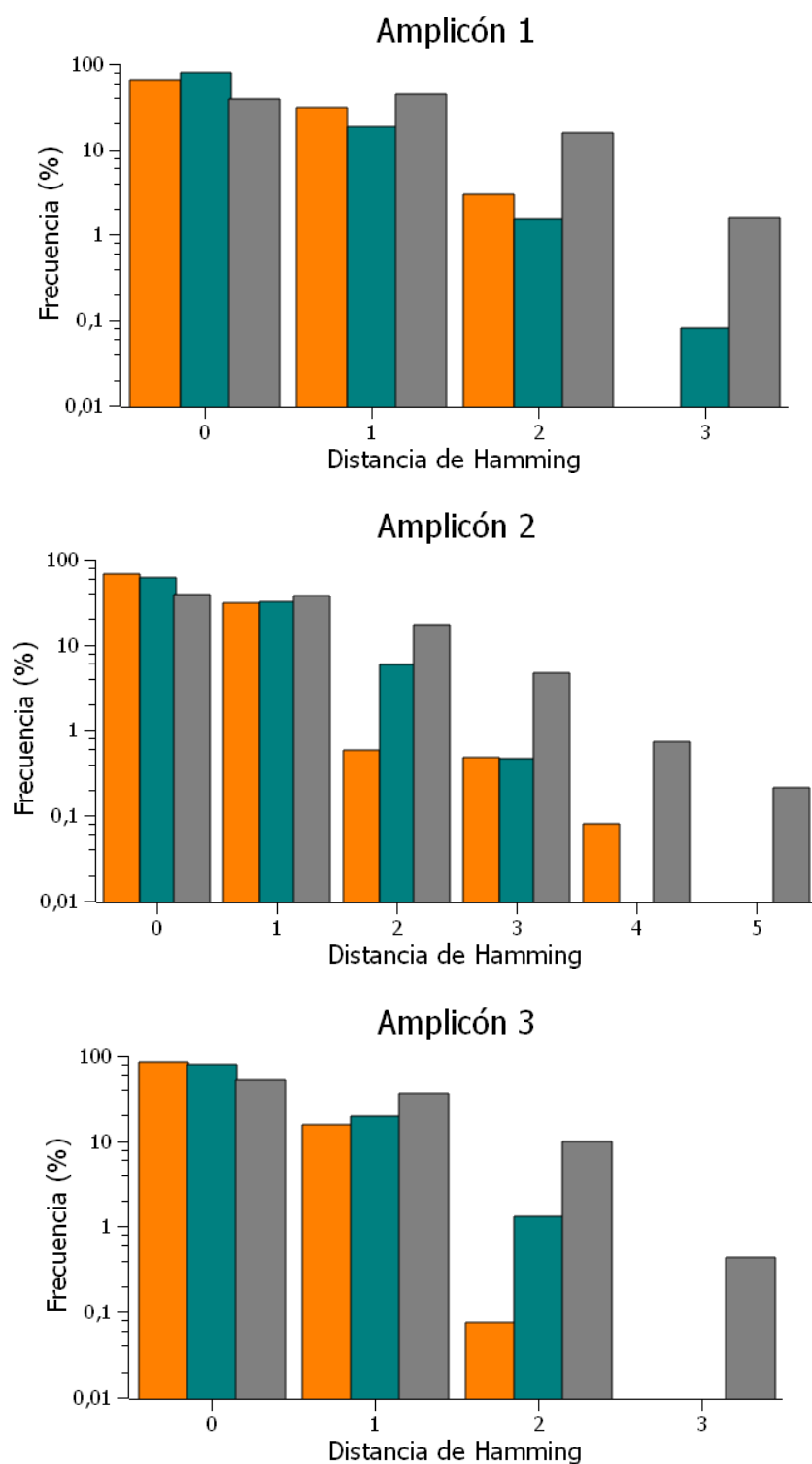


Figura 53: Frecuencia de haplotipos en función de su distancia de Hamming respecto a la secuencia consenso en el pase número 60 de la evolución. Poblaciones C30_{p2.1} (●), Ap2.2 (●) y C43_{p2.3} (●).

La variación a lo largo de la adaptación de las frecuencias de las mutaciones representadas en la población ancestral, Q β _{p2}, por encima del 0.5 % se muestra en la Tabla 47. De estas mutaciones, ninguna llega a alcanzar el valor del 5 % en la población C30_{p2.1}, mientras que en C43_{p2.3} lo hacen 3 y en Ap2.2 solo una (siendo esta una de las que lo estaba también en C43_{p2.3}).

		Frecuencia (%)			
	Mutación	Q β _{P2}	C30 _{P2.1}	A _{P2.2}	C43 _{P2.3}
Amplicón 1	U1295G	0.67	0.09	0.19	3.17
	G1312A	0.70	0.32	6.54	91.45
	U1328G	1.19	0.62	0.52	0.58
Amplicón 2	A2187C	1.00	0.80	4.17	62.02
	C2189U	1.27	0.00	0.00	0.00
	G2211A	0.68	0.56	0.47	1.24
	A2212U	0.68	0.56	0.47	1.24
	U2213C	0.68	0.68	0.47	2.02
	C2391A	0.57	0.25	0.23	0.29
Amplicón 3	C3545U	0.67	1.49	1.60	8.82
	A3603G	0.50	0.13	0.54	0.59

Tabla 47: Frecuencias alcanzadas a pase número 60 por las mutaciones representadas en la población ancestral en proporción mayor o igual al 0.5 % en las poblaciones C30_{P2.1}, A_{P2.2} y C43_{P2.3}. Las celdas sombreadas en gris indican las mutaciones que alcanzan una representación por encima del 5 %.

A continuación se procedió a identificar las mutaciones que no estaban presentes por encima del 0.5 % en la población ancestral y que alcanzaron una proporción mayor del 5 % a pase número 60 en algunas de las poblaciones evolucionadas. Estas mutaciones se muestran en la Tabla 48, junto con sus frecuencias en cada pase.

		Frecuencia (%)			
	Mutación	Q β _{P2}	C30 _{P2.1}	A _{P2.2}	C43 _{P2.3}
Amplicón 1	A1088G	0.21	0.19	99.01	51.65
	G1122A	0.00	75.00	0.33	0.15
	U1295C	0.06	2.65	4.01	11.40
Amplicón 2	C2210U	0.00	13.71	0.08	0.09
	G2217A	0.25	0.06	0.00	8.51
	C2245U	0.00	7.00	0.00	0.00
	C2405U	0.00	0.13	12.45	0.38
	U2450C	0.00	0.10	0.00	14.16
	C2452U	0.00	0.09	95.88	13.11
Amplicón 3	U3402C	0.20	0.16	12.17	89.16
	G3494A	0.00	0.20	0.08	25.72

Tabla 48: Mutaciones que no estaban presentes en la población ancestral en más del 0.5 % de las lecturas y que alcanzan una representación mayor o igual al 5 % en alguna de las poblaciones evolucionadas. Las celdas sombreadas en gris indican los casos en los que se supera el 5 %.

Nuevamente se observa que en la línea C30_{P2.1} el número de mutaciones que alcanzan la frecuencia del 5 % es muy bajo (3 mutaciones), mientras que ese número aumenta sensiblemente en la línea C43_{P2.3} (7 mutaciones) y se mantiene en valores intermedios

en la línea $A_{P2.2}$ (4 mutaciones). De estas mutaciones, no hay ninguna coincidente entre $C30_{P2.1}$ y cualquiera de las otras dos, mientras que hay 3 que coinciden entre $C43_{P2.3}$ y $A_{P2.2}$.

Los valores de los índices de diversidad basados en incidencia ($F_{mín}$ y $dHpl$) se muestran en la Tabla 49 y en la Fig. 54. En ambos casos también se incluyen los valores de la población ancestral $Q\beta_{P2}$.

		N	Hpl	$dHpl(\cdot 10^{-4})$	nM	$F_{mín}(\cdot 10^{-7})$
Amplicón 1	$Q\beta_{P2}$	266858	30	1.12	28	3.86
	$C30_{P2.1}$ (P60)	188441	76	4.03	65	12.68
	$A_{P2.2}$ (P60)	206263	74	3.59	71	12.65
	$C43_{P2.3}$ (P60)	265606	64	2.41	32	4.43
Amplicón 2	$Q\beta_{P2}$	294855	49	1.66	50	5.15
	$C30_{P2.1}$ (P60)	222745	77	3.46	71	9.69
	$A_{P2.2}$ (P60)	189954	49	2.58	45	7.20
	$C43_{P2.3}$ (P60)	252206	78	3.09	44	5.30
Amplicón 3	$Q\beta_{P2}$	223334	39	1.75	38	6.19
	$C30_{P2.1}$ (P60)	245275	74	3.02	72	10.67
	$A_{P2.2}$ (P60)	242689	62	2.55	59	8.84
	$C43_{P2.3}$ (P60)	300033	58	1.93	38	4.61

Tabla 49: Valores de los índices de diversidad basados en incidencia en el pase número 60 de las líneas evolutivas $C30_{P2.1}$, $A_{P2.2}$ y $C43_{P2.3}$ y su correspondiente ancestro $Q\beta_{P2}$. La definición exacta de los parámetros mostrados en la tabla se encuentra en el apartado 4.2.7. de Materiales y Métodos.

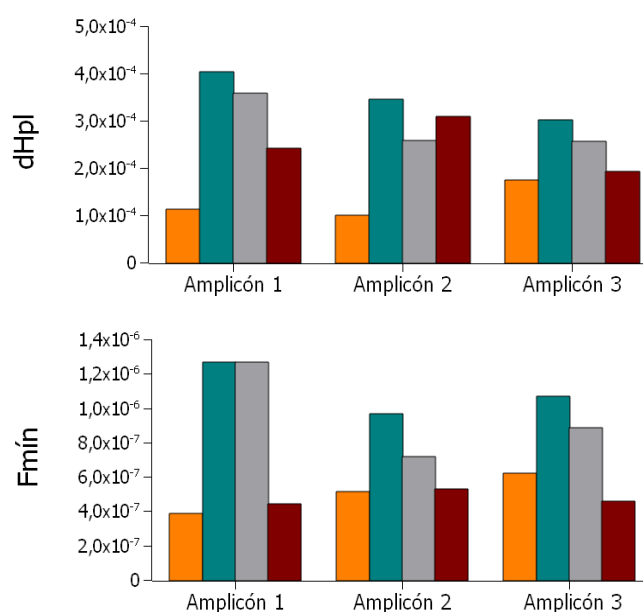


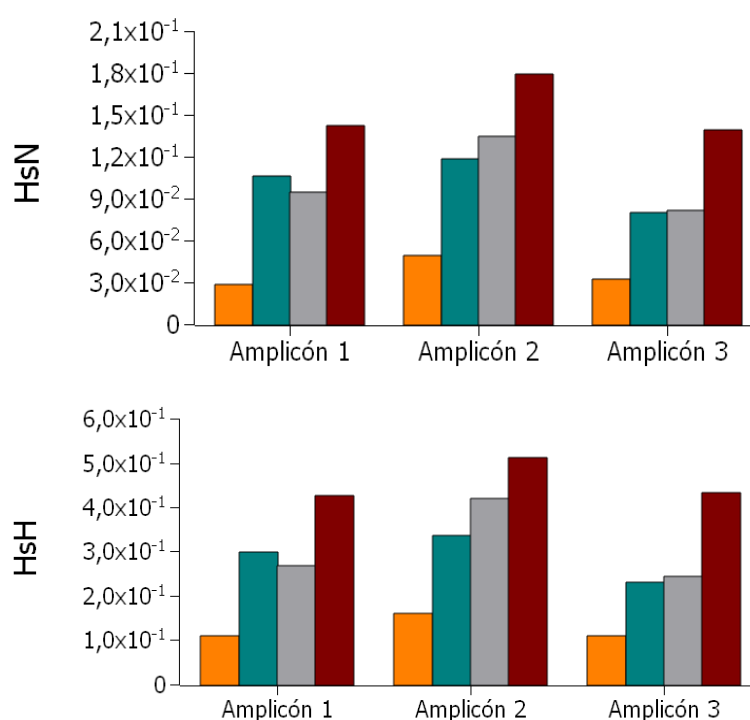
Figura 54: Valores de la densidad de haplotipos y frecuencia de mutación mínima en la población ancestral y en el pase número 60 de las líneas evolucionadas. Población ancestral $Q\beta_{P2}$ (●), $C30_{P2.1}$ (●), $A_{P2.2}$ (●) y $C43_{P2.3}$ (●).

En todos los casos los valores de la densidad de haplotipos son mayores en las poblaciones evolucionadas que en la población ancestral, aunque no se observa una pauta que se mantenga para las tres líneas evolutivas en los tres amplicones. Los valores de $F_{mín}$ aumentan claramente en el caso de C30_{P2.1} y A_{P2.2} y no lo hacen (o lo hacen muy ligeramente) en el caso de C43_{P2.3}.

Los valores de los índices de diversidad basados en abundancia se muestran en la Tabla 50 y en la Fig. 55.

		N	Hs	HsN($\cdot 10^{-2}$)	HsH	π ($\cdot 10^{-3}$)	Fmáx($\cdot 10^{-4}$)
Amplicón 1	Q β_{P2}	266858	0.36	2.88	0.11	0.39	1.98
	C30 _{P2.1} (P60)	188441	1.29	10.64	0.30	2.22	13.41
	A _{P2.2} (P60)	206263	1.16	9.45	0.27	1.56	8.03
	C43 _{P2.3} (P60)	265606	1.78	14.22	0.43	3.99	29.46
Amplicón 2	Q β_{P2}	294855	0.62	4.91	0.16	0.64	3.20
	C30 _{P2.1} (P60)	222745	1.462	11.88	0.34	1.93	10.38
	A _{P2.2} (P60)	189954	1.63	13.43	0.42	2.59	13.71
	C43 _{P2.3} (P60)	252206	2.23	17.93	0.51	4.35	27.57
Amplicón 3	Q β_{P2}	223334	0.40	3.26	0.11	0.40	2.02
	C30 _{P2.1} (P60)	245275	0.99	8.02	0.23	1.15	5.84
	A _{P2.2} (P60)	242689	1.01	8.15	0.24	1.49	8.02
	C43 _{P2.3} (P60)	300033	1.76	13.95	0.43	3.59	21.14

Tabla 50: Valores de los índices de diversidad basados en abundancia en el pase número 60 de las líneas evolutivas C30_{P2.1}, A_{P2.2} y C43_{P2.3} y su correspondiente ancestro Q β_{P2} . La definición exacta de los parámetros mostrados en la tabla se encuentra en el apartado 4.2.7. de Materiales y Métodos.



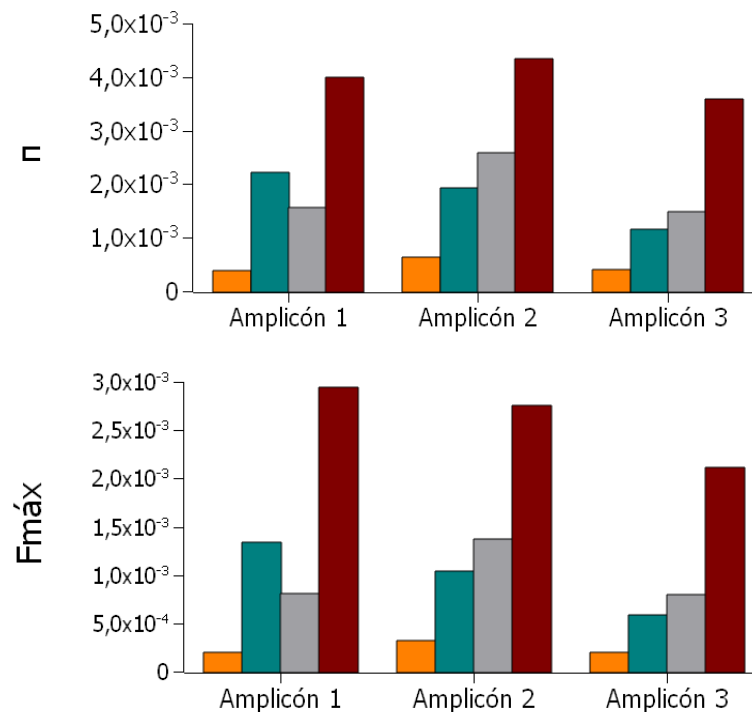


Figura 55: Valores de los índices de diversidad basados en abundancia en la población ancestral y en el pase número 60 de las líneas evolucionadas. Población ancestral Qβ_{p2} (●), C30_{p2.1} (●), Ap_{2.2} (●) y C43_{p2.3} (●).

Para todos los parámetros, amplicones y líneas evolutivas se observan aumentos claros respecto a la población ancestral. En todos los casos, la línea evolutiva que alcanza valores mayores es la C43_{p2.3}. Las líneas C30_{p2.1} y Ap_{2.2} presentan menores diferencias entre ellas, unas veces a favor de una y otras de la otra.

5.7.4 Variación de los índices de diversidad en función de la temperatura de evolución

La clara diferencia observada entre los valores de los índices de diversidad, sobre todo los basados en abundancia, de las poblaciones C43_{p2.3} y C30_{p2.1} planteó la cuestión de si habría una relación entre la diversidad generada en una población y su temperatura de evolución. Para dar respuesta a esta pregunta, se hizo uso de varias poblaciones que habían sido propagadas al mismo tiempo que las líneas C30_{p2}, a las temperaturas constantes de 33 °C, 37 °C y 40 °C utilizando la misma población ancestral, Qβ_{p2}, durante 60 pases seriados en triplicado (Fig. 11). Hay que señalar que una de las líneas propagadas a 33 °C se perdió durante la realización del experimento, por lo que para esta temperatura solo tenemos duplicados. A este conjunto de líneas evolutivas, se añadieron también las C43_{p2}, de modo que los parámetros a analizar se pudieran comparar en poblaciones evolucionadas a dos temperaturas por debajo de la óptima (30 °C y 33 °C), a la óptima (37 °C) y a dos por encima de la óptima (40 °C y 43 °C). El

efecto de todas estas temperaturas en la replicación de Q β (población Q β _{P2}) se muestra en la Fig. 56, en la que puede observarse que las temperaturas inferiores a la óptima apenas causan diferencias en el *fitness* del virus, mientras que las superiores causan un claro descenso.

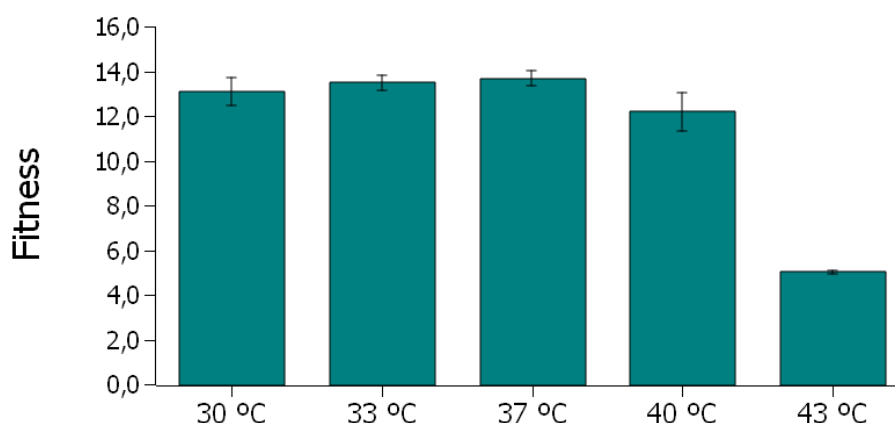


Figura 56: *Fitness* de la población ancestral Q β _{P2} a diferentes temperaturas. Cada barra corresponde a la media de tres determinaciones realizadas en triplicado.

A continuación se determinó el *fitness* de todas las líneas evolutivas a la misma temperatura a la que habían sido propagadas de forma constante. Los valores obtenidos se refirieron al obtenido para la población ancestral Q β _{P2} (Fig. 57).

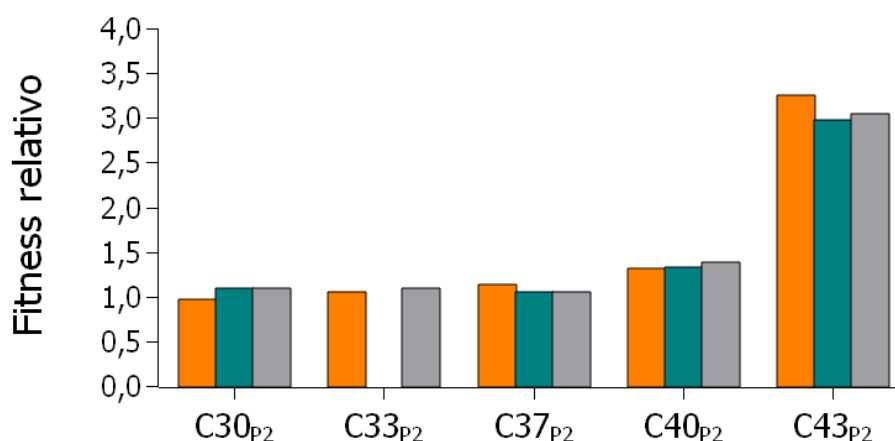


Figura 57: Valores de *fitness* de las líneas evolutivas C30_{P2}, C33_{P2}, C37_{P2}, C40_{P2} y C43_{P2} a pase número 60. Los ensayos de *fitness* se realizaron a la temperatura de evolución correspondiente en cada una de las líneas. (●) réplica 1, (●) réplica 2, (●) réplica 3. El virus de referencia utilizado fue la población ancestral Q β _{P2} que es el origen de todas las líneas evolutivas representadas.

Se observa que las líneas propagadas a 30 °C, 33 °C y 37 °C apenas experimentan ganancias de *fitness*. A 40 °C hay ganancias moderadas en todas las poblaciones y a 43 °C hay claros aumentos. Estos resultados indican que las presiones selectivas más intensas son las que producen mayores ganancias de *fitness*.

Las mutaciones fijadas u observables como polimorfismos en la secuencia consenso se muestran en la Tabla 51. Cuando se analizan las mutaciones coincidentes con las presentes en algunas de las líneas C43_{P2} (Tabla 45), encontramos 2 mutaciones en las líneas C30_{P2} (C3545U y G3945A), 1 en las líneas C33_{P2} (G3945A), 3 en las líneas C37_{P2} (U1295C/G, A2187C, G3945A) y 5 en las líneas C40_{P2} (G1312A, U2006A, U2776C, U3402C y G3945A).

	U458C	C562A	U574G	U750A	C900U	U1025C	A1088G	C1228U	U1295C/G	G1312A	G1371A
C33 _{P2}	X	X	X	X	X	X	X	X	X	X	X
C37 _{P2}											
C40 _{P2}											

	A1569G	C1806U	A1823G	A1915G	A1930G	U2006A	C2011A	U2016G	A2187C	C2201U	C2210U
C33 _{P2}	X	X	X	X	X	X	X	X	X	X	X
C37 _{P2}											
C40 _{P2}											

	G2217A	G2223A	C2228U	A2739G	G2741A	U2776C	U2785G	U3014C	C3065U	U3402C	G3945A
C33 _{P2}	X	X	X	X	X	X	X	X	X	X	X
C37 _{P2}											
C40 _{P2}											

wt	mut<wt	mut=wt	mut>wt	mut
----	--------	--------	--------	-----

Tabla 51: Mutaciones y polimorfismos en las secuencias consenso de las líneas evolutivas C33_{P2}, C37_{P2} y C40_{P2} en el pase número 60 de la adaptación. La primera fila muestra los diferentes cambios nucleotídicos observados. El estado de las mutaciones se muestra con un sombreado de diferente intensidad (ver parte inferior de la tabla) en los cuadrados situados debajo de cada mutación (de izquierda a derecha, réplicas 1, 2 y 3 respectivamente). La posición 1295 experimentó cambios U/C o U/G que, con frecuencia, coexistían en la misma población. En la figura no se ha hecho distinción entre ambos tipos de cambio. El recuadro correspondiente a la línea C33_{P2.2} (que se perdió) aparece con signo X.

Para comparar la estructura de las poblaciones obtenidas a pase número 60, se realizó un análisis mediante NGS de la región genómica 1 para una línea representativa de cada conjunto. Los procedimientos y umbral de representación de las secuencias utilizadas en el análisis fueron los mismos que en el caso de las poblaciones Qβ_{P2} y Qβ_{P25} (ver apartados 4.2.6., 4.2.7. y 5.1.1.). El número de lecturas totales, las obtenidas tras el procesamiento bioinformático y las correspondientes a las secuencias representadas en más del 0.05 % del total de lecturas se muestran en la Tabla 6.

El número de mutaciones que se encuentran representadas por encima del 5 % en las diferentes líneas analizadas se muestra en la Tabla 52 junto con las frecuencias alcanzadas por esas mismas mutaciones en la población ancestral. Solamente dos de estas mutaciones U1295G y G1312A estaban presentes en la población ancestral en una frecuencia mayor del 0.5 %.

	Mutación	Frecuencia (%)					
		Q β _{P2}	C30 _{P2.1}	C33 _{P2.1}	C37 _{P2.3}	C40 _{P2.1}	C43 _{P2.3}
Amplícón 1	A1088G	0.21	0.19	0.15	0.27	2.97	51.65
	G1122A	0.00	75.00	0.27	0.12	0.44	0.15
	C1137U	0.00	0.19	8.73	0.42	0.17	0.15
	U1295C	0.06	2.65	10.72	53.42	5.23	11.40
	U1295G	0.67	0.09	0.14	0.39	5.75	3.17
	G1312A	0.70	0.32	0.38	15.46	43.77	91.45

Tabla 52: Mutaciones que alcanzan una representación mayor o igual al 5 % a pase número 60 de la evolución de las poblaciones C30_{P2.1}, C33_{P2.1}, C37_{P2.3}, C40_{P2.1} y C43_{P2.3}. Las celdas sombreadas en gris indican los casos en los que se supera el 5 %.

A continuación se representaron las distribuciones de frecuencias de los haplotipos presentes en más del 0.5 % de las lecturas en función de su número de orden (Fig. 58) y de su distancia de Hamming a la secuencia consenso (Fig. 59), no observándose en el primer caso diferencias relevantes entre las poblaciones analizadas.

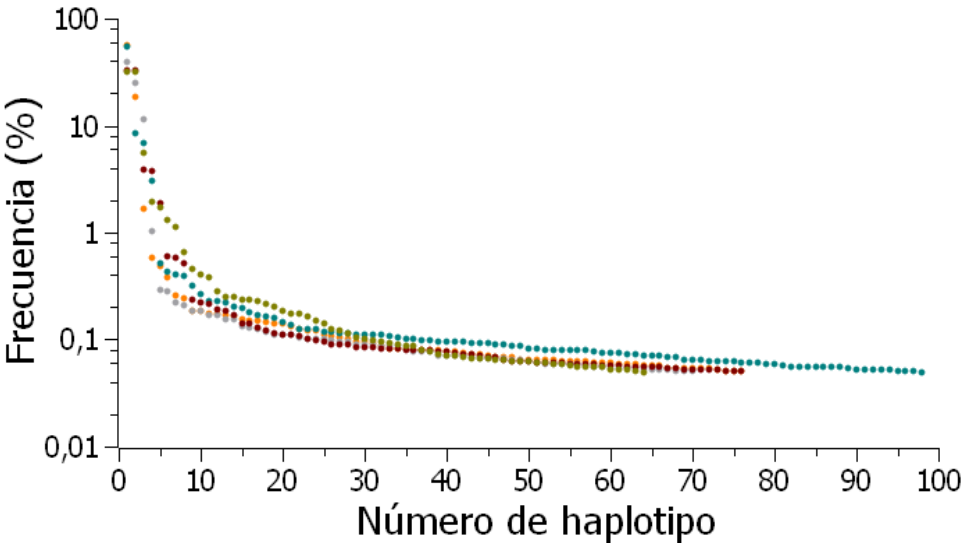


Figura 58: Frecuencias de los haplotipos representados en más del 0.05 % de las lecturas obtenidas en el amplícón 1 a pase número 60. Poblaciones C30_{P2.1} (●), C33_{P2.2} (●), C37_{P2.3} (●), C40_{P2.1} (●) y C43_{P2.3} (●). El número de cada haplotipo representa el número de orden en relación a su frecuencia.

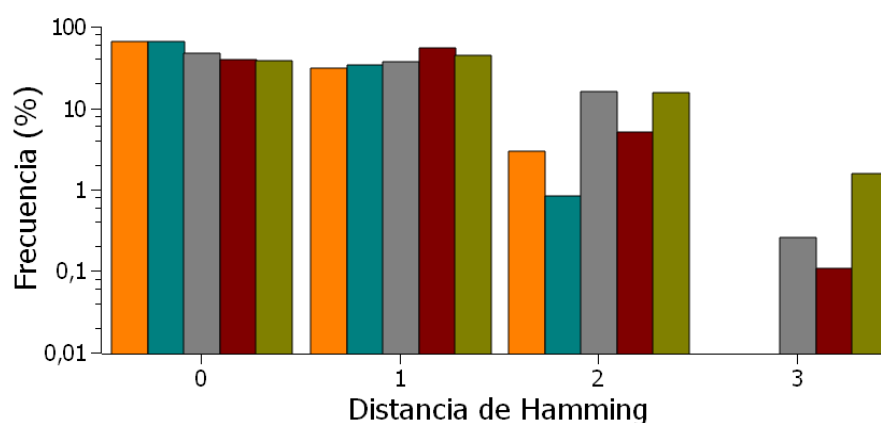


Figura 59: Frecuencia de haplotipos en función de su distancia de Hamming respecto a la secuencia consenso obtenida a pase número 60 en el amplicón 1. Poblaciones C30_{P2.1} (●), C33_{P2.1} (●), C37_{P2.3} (●), C40_{P2.1} (●) y C43_{P2.3} (●).

Se observa que el haplotipo a distancia de Hamming 0 es más frecuente en las poblaciones que han evolucionado a las temperaturas más baja (30 °C y 33 °C), mostrando un descenso en su representación en las líneas propagadas 37 °C, 40 °C y 43 °C. Salvo algunas excepciones, los haplotipos a distancias de Hamming igual o superior a 1 están más representados en las poblaciones que han evolucionado a 37 °C, 40 °C o 43 °C que en las que lo han hecho a 30 °C o 33 °C. Estos resultados son compatibles con la generación de poblaciones más diversas a temperaturas más altas.

Por último, determinamos los valores de los índices de diversidad para todas estas poblaciones (Tablas 53 y 54 y Figs. 60 y 61).

		N	Hpl	dHpl($\cdot 10^{-4}$)	nM	Fmín($\cdot 10^{-7}$)
Amplicón 1	Q β _{P2}	266858	30	1.12	28	3.86
	C30 _{P2.1} (P60)	188441	76	4.03	65	12.68
	C33 _{P2.1} (P60)	184375	98	5.31	93	18.54
	C37 _{P2.3} (P60)	200913	71	3.53	52	9.51
	C40 _{P2.1} (P60)	171155	76	4.44	47	10.10
	C43 _{P2.3} (P60)	265606	64	2.41	32	4.43

Tabla 53: Valores de los índices de diversidad basados en incidencia de las líneas evolutivas C30_{P2.1}, C33_{P2.1}, C37_{P2.3}, C40_{P2.1} y C43_{P2.3} obtenidas a pase número 60 y su correspondiente ancestro Q β _{P2}. La definición exacta de los parámetros mostrados en la tabla se encuentra en el apartado 4.2.7. de Materiales y Métodos.

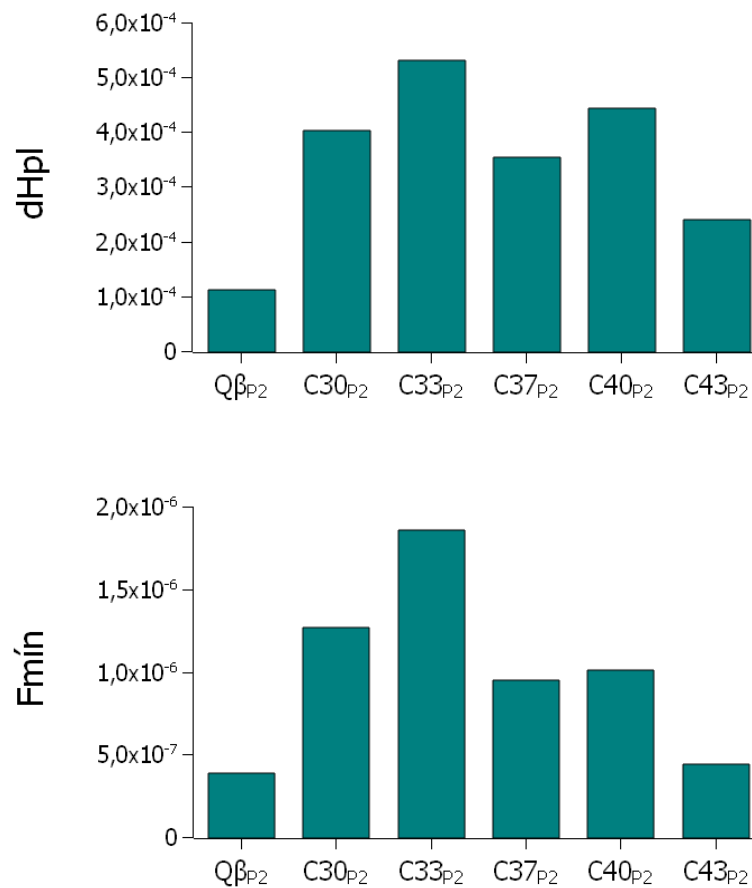


Figura 60: Densidad de haplotipos y frecuencia de mutación mínima para las líneas evolutivas $C30_{P2.1}$, $C33_{P2.1}$, $C37_{P2.3}$, $C40_{P2.1}$ y $C43_{P2.3}$ (pase 60) y su correspondiente ancestro $Q\beta_{P2}$.

La observación más clara es que todas las poblaciones evolucionadas presentan valores mayores para $F_{mín}$ y $dHpl$ que la población ancestral, con la única excepción de la línea $C43_{P2.3}$ que presenta un valor muy parecido. Los valores mayores para ambos índices se obtienen a las temperaturas inferiores a la óptima (30 °C y 33 °C).

		N	Hs	HsN($\cdot 10^{-2}$)	HsH	π ($\cdot 10^{-3}$)	Fmáx($\cdot 10^{-4}$)
Amplificón 1	$Q\beta_{P2}$	266858	0.36	2.88	0.11	0.39	1.98
	$C30_{P2.1}$ (P60)	188441	1.29	10.65	0.30	2.22	13.41
	$C33_{P2.1}$ (P60)	184375	1.63	13.46	0.36	2.42	12.90
	$C37_{P2.3}$ (P60)	200913	1.56	12.75	0.37	3.34	25.48
	$C40_{P2.1}$ (P60)	171155	1.68	13.97	0.39	3.33	24.07
	$C43_{P2.3}$ (P60)	265606	1.78	14.22	0.43	3.99	29.46

Tabla 54: Valores de los índices de diversidad basados en abundancia de las líneas evolutivas $C30_{P2.1}$, $C33_{P2.1}$, $C37_{P2.3}$, $C40_{P2.1}$ y $C43_{P2.3}$ obtenidas a pase número 60 y su correspondiente ancestro $Q\beta_{P2}$. La definición exacta de los parámetros mostrados en la tabla se encuentra en el apartado 4.2.7. de Materiales y Métodos.

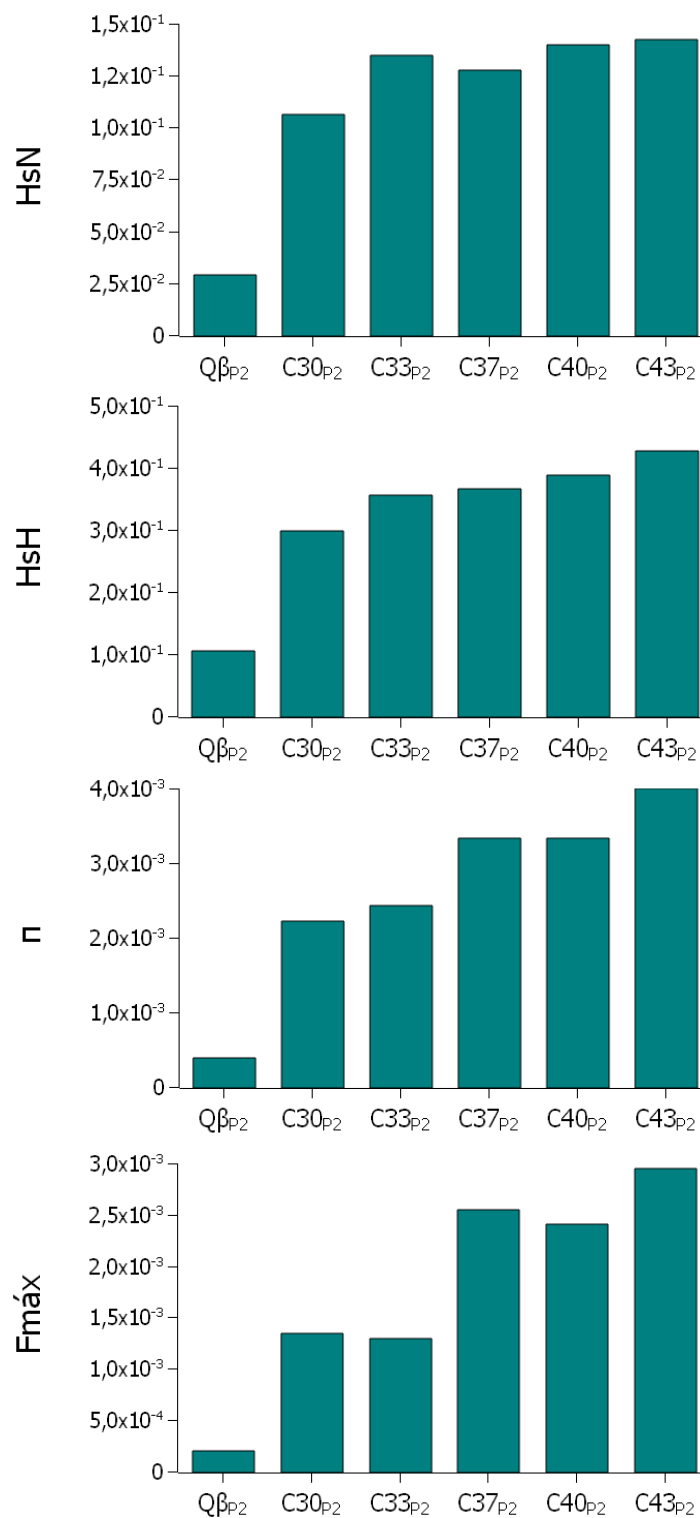


Figura 61: Valores de los índices de diversidad basados en abundancia para las líneas evolutivas C30P2.1, C33P2.1, C37P2.3, C40P2.1 y C43P2.3 (pase 60) y su correspondiente ancestro QβP2.

Se observa que la propagación a cualquier temperatura produce un claro aumento de los valores de los índices de diversidad basados en abundancia, siendo estos mayores cuanto mayor es la temperatura de evolución.

6. DISCUSIÓN

En esta tesis doctoral se ha planteado un experimento de evolución en el que el bacteriófago Q β ha sido propagado a temperaturas distintas de la óptima, utilizando varias pautas de cambio y partiendo de diferentes poblaciones ancestrales. Se ha estudiado la influencia de estos factores en la velocidad de adaptación (mediante determinaciones de la tasa de crecimiento en distintos puntos del proceso) y en las vías adaptativas seguidas (mediante la identificación de las mutaciones aparecidas en la secuencia consenso). Con objeto de analizar en profundidad los cambios producidos en los espectros de mutantes de las distintas poblaciones se ha realizado un análisis mediante secuenciación masiva de tres amplicones, localizados en regiones genómicas en las que, por los resultados de experimentos previos, sabemos que se seleccionan mutaciones durante la evolución a temperatura elevada (Arribas y col., 2014; 2016; 2018; Kashiwagi y col., 2014). Los valores de la densidad de haplotipos y la frecuencia de mutación mínima han permitido estimar la diversidad genética total contenida en las poblaciones, independientemente de cómo de abundante sea cada mutación o cada haplotipo. Las distribuciones de frecuencias de haplotipos en función de su número de orden o de su distancia de Hamming, así como los valores de la entropía de Shannon, el índice π o la frecuencia de mutación máxima respecto a la secuencia consenso han sido utilizados para analizar la distribución de mutaciones y haplotipos, por lo que informan sobre la estructura interna y el grado de complejidad de las poblaciones. A continuación se discuten los resultados más relevantes obtenidos.

6.1 La propagación de un clon biológico de Q β en condiciones estándar aumenta la diversidad y complejidad de la población

Para analizar el efecto de la diversidad genética preexistente y de la estructura de la población en la adaptación de Q β a temperatura elevada, necesitábamos en primer lugar disponer de poblaciones que difirieran en estos parámetros. Para ello se partió de una placa de lisis generada a partir de la expresión del plásmido pBRT7Q β (virus Q β_{WT}). Puesto que en las condiciones del plaqueo utilizadas, existe una alta probabilidad de que las placas de lisis se generen a partir de la replicación de un único genoma viral durante un número reducido de generaciones, es de suponer el virus Q β_{WT} corresponda a una población bastante homogénea. A pesar de esto, se ha demostrado que, dada la alta tasa de error de los virus RNA (Batschelet y col., 1976; Domingo y col., 1997; Drake y Holland, 1999; Sanjuán y col., 2010) y de Q β en particular (Bradwell y col., 2013), incluso las placas de lisis pueden contener cierta heterogeneidad, como así se ha comprobado en experimentos en los que se han analizado los componentes subclonales de placas de lisis aisladas (Duarte y col., 1994; Cervera y Elena, 2016) o se ha estudiado la acumulación de mutaciones en experimentos de propagación de virus a través de pases placa a placa (Chao, 1990; Clarke y col., 1993; Yuste y col., 2005; Lázaro y col., 2003; Escarmís y col., 2008).

La propagación del virus $Q\beta_{WT}$ en las condiciones estándar de nuestro laboratorio, utilizando tamaños poblacionales elevados, para favorecer la generación de mutaciones beneficiosas que no sean eliminadas por deriva génica, es esperable que produzca importantes cambios tanto en la estructura de la población como en su grado de heterogeneidad. Durante la replicación del virus se genera un conjunto diverso de mutantes que serán expuestos a la acción de la selección natural, de modo que los que tengan efectos negativos en las condiciones ambientales ensayadas serán eliminados o mantenidos en bajas frecuencias, mientras que los que sean beneficiosos aumentarán su representación en la población. La alta tasa de error de la replicación viral impide que el genoma mejor adaptado domine completamente la población, la cual estará compuesta por un espectro de mutantes cuya estructura y diversidad variará en función del número de rondas de replicación (o pases) experimentados (Domingo y col., 2012; Andino y Domingo, 2015; Domingo y Perales, 2019). La teoría de cuasiespecies propone que cuando una población se replica con alta tasa de error, originando mutantes que se diferencian en sus valores de *fitness*, dado un tiempo suficientemente largo llegará a un estado estacionario en el que su estructura permanecerá constante (Biebricher y Eigen, 2006).

En esta tesis doctoral, hemos propagado el virus $Q\beta_{WT}$ durante 2 y 25 pases seriados, un número de pases que, aunque probablemente no es suficiente para alcanzar el estado estacionario, sí lo habrá sido para aumentar la diversidad genética y el grado de complejidad de las poblaciones $Q\beta_{P2}$ y $Q\beta_{P25}$ obtenidas. Algunos estudios previos han demostrado que la propagación de varios clones del virus de la inmunodeficiencia humana, a lo largo de diez pases con tamaños poblacionales elevados, conducen a aumentos de *fitness* y de la heterogeneidad de la cuasiespecies (Bordería y col., 2010). También, en otro estudio, llevado a cabo con el virus de la hepatitis C, se demostró que, después de 200 pases realizados en condiciones constantes, la población continuaba con su proceso de diversificación (Gallego y col., 2020).

El análisis mediante NGS de tres amplicones de $Q\beta_{P2}$ y $Q\beta_{P25}$ muestra interesantes diferencias entre las dos poblaciones. En primer lugar, la población $Q\beta_{P25}$ presenta mayor densidad de haplotipos diferentes que $Q\beta_{P2}$, lo cual indica que a lo largo de los pases se ha aumentado la diversidad genética.

La representación de las frecuencias de los distintos haplotipos contenidos en las dos poblaciones en función de su número de orden (Fig. 17) indica que la población $Q\beta_{P2}$ está dominada por un haplotipo mayoritario (que en todos los casos representa más del 90 % de las lecturas) cuya secuencia no contiene mutaciones respecto al virus $Q\beta_{WT}$. Los haplotipos siguientes al mayoritario en esta población escasamente alcanzan el 1 % (Tabla 11). En el caso de la población $Q\beta_{P25}$ el haplotipo mayoritario está presente en frecuencias menores, sobre todo en el caso del amplicón 2, donde está en una proporción del 55 %, coexistiendo con otros que están en frecuencias del 15 % o el 8 %, lo cual supone un fuerte cambio en su estructura. Estas diferencias se ven confirmadas en la representación de las frecuencias de haplotipos en función de su distancia de Hamming a la secuencia consenso (Figs. 18 y 19).

Los valores de los índices de diversidad basados en abundancia también son en todos los casos mayores para $Q\beta_{P25}$ que para $Q\beta_{P2}$ y alcanzan los máximos valores en el amplicón 2, indicando que de las tres regiones analizadas por NGS, la correspondiente a este amplicón es la que más cambios experimenta en su estructura durante la propagación del virus a 37 °C.

6.2 Durante la propagación de $Q\beta$ a 37 °C se genera un espectro de mutantes que contiene en bajas frecuencias mutaciones relevantes para la adaptación a 43 °C

Cuando se analizaron las mutaciones que están representadas por encima del 0.5 % en las dos poblaciones analizadas se obtuvo un resultado inesperado. Ya, a pase número 2, observamos que 3 mutaciones de un total de 10 habían sido previamente identificadas en otros experimentos de adaptación a temperatura elevada llevados a cabo con $Q\beta$. En la población $Q\beta_{P25}$ ese número aumentó a 8 de 17 mutaciones totales (Arribas y col., 2014; 2016; 2018; Kashiwagi y col., 2014; Lázaro y col.; 2018). Algunas de esas mutaciones (U1295C, A2187C) también habían sido detectadas en experimentos de adaptación a 37 °C (Arribas y col., 2014; 2016), pero el resto nunca habían sido identificadas en la secuencia consenso de poblaciones de $Q\beta$ propagadas a esta temperatura. Estos resultados sugieren la interesante posibilidad de que durante las primeras rondas de replicación, que permiten generar una población diversa a partir de una población clonal, se genera un conjunto de mutaciones que permanecen en frecuencias bajas en las condiciones estándar de propagación del virus. Estas mutaciones podrían hacerse mayoritarias cuando se produce un cambio de ambiente, en este caso el aumento de la temperatura de replicación. Algunas de las mutaciones que estaban por encima del 0.5 % en $Q\beta_{P2}$ están ausentes en $Q\beta_{P25}$, lo cual podría indicar la acción de la selección negativa sobre ellas. Hay otras cuya frecuencia aumenta entre los pases 2 y 25 y otras que solo alcanzan el umbral del 0.5 % en $Q\beta_{P25}$. Estas mutaciones se localizan preferentemente en la región correspondiente al amplicón 2, sugiriendo que de las tres regiones genómicas analizadas, es en esta donde la acción de la selección natural es más potente durante la replicación a 37 °C.

6.3 La diversidad genética generada durante la evolución a 37 °C acelera la adaptación a 43 °C

La propagación a 43 °C de las poblaciones ancestrales $Q\beta_{P2}$ y $Q\beta_{P25}$ muestra que en ambos casos se obtienen aumentos notables de *fitness* que, después de 60 pases, no difieren significativamente en función de la población ancestral utilizada (Fig. 25). Sin embargo, la dinámica de la ganancia de *fitness* a 43 °C es claramente diferente. Cuando se parte de la población más heterogénea, $Q\beta_{P25}$, las ganancias son mucho más rápidas que cuando se parte de la población $Q\beta_{P2}$, lo cual podría indicar que algunos de los

genomas que existen en el espectro de mutantes de $Q\beta_{P25}$ contienen ya mutaciones con efectos beneficiosos a 43 °C o que favorecen la adaptación a esta temperatura. Los resultados estarían de acuerdo con que, en $C43_{P25}$, el primer evento de fijación de mutaciones ocurre a pases más tempranos (pase 10 o anterior) que en $C43_{P2}$ e implica la coincidencia de varias mutaciones (A1088G, A2222C y U3311G) que podrían estar ya presentes en algún genoma minoritario de la población $Q\beta_{P25}$. Esas 3 mutaciones también están entre las que se fijan primero en una de las líneas $C43_{P2}$, pero lo hacen más tarde, sugiriendo que ese genoma minoritario también podría estar presente en la población ancestral $Q\beta_{P2}$ aunque en menor proporción que en $Q\beta_{P25}$.

El análisis de las secuencias consenso a lo largo de la adaptación muestra que la mutación más representada en el amplicón 2 de $Q\beta_{P25}$ (A2187C) está ausente de las secuencias consenso de las tres líneas evolutivas obtenidas tras la adaptación a 43 °C a partir de este ancestro. La selección en contra de esta mutación es muy rápida, ya que a pase 10 es indetectable. El hecho de que la desaparición de A2187C coincida con el primer evento de fijación sugiere que A2187C está localizada en genomas que contienen una combinación de mutaciones que es menos favorable a 43 °C que la que alcanza la fijación en primer lugar. Otra posibilidad es que A2187C tenga efectos negativos a 43 °C. Sin embargo, el hecho de que esta misma mutación se haya seleccionado en experimentos previos de adaptación a 43 °C (Arribas y col., 2014; 2016) y aparezca como polimorfismo de alta intensidad en todas las líneas evolutivas obtenidas a partir de $Q\beta_{P2}$ estaría en contra de esa posibilidad. También podría suceder que durante los pases previos a la detección de A2187C en $C43_{P2}$ se seleccionaran otras mutaciones que proporcionarían un contexto más favorable para su permanencia.

Cuando se analizan las mutaciones, fijadas o polimórficas, que están presentes en las secuencias consenso a pase número 60, se observa que hay un conjunto de mutaciones que se seleccionan mayoritariamente durante la propagación a 43 °C, independientemente de cual sea la población ancestral (Fig. 28). Estas mutaciones suelen ser no sinónimas y coinciden con otras identificadas en experimentos previos de adaptación de $Q\beta$ a temperatura elevada (Arribas y col., 2014; 2016; 2018; Kashiwagi y col., 2014; Lázaro y col.; 2018), lo cual es un claro indicador de su valor selectivo. Algunas de ellas, además, estaban ya representadas por encima del 0.5 % en al menos una de las poblaciones ancestrales (A1088G, U1295G/C y G1312A).

Las mutaciones exclusivas de los conjuntos de líneas $C43_{P2}$ o $C43_{P25}$ suelen estar representadas en una única línea, lo que sugiere que probablemente son menos relevantes para la adaptación. Estas mutaciones, además, contienen una mayor proporción de cambios sinónimos.

La mayor convergencia evolutiva observada entre las líneas $C43_{P25}$ que entre las líneas $C43_{P2}$ podría ser debida a que la adaptación en el primer caso tiene lugar a partir de al menos una parte de la diversidad genética preexistente en la población ancestral $Q\beta_{P25}$. Estos resultados están de acuerdo con otros estudios cuantitativos (Hill, 1982; Roff, 1997) que predicen que a corto plazo la selección natural actúa preferentemente

sobre la diversidad genética preexistente (Quakkelaar y col., 2007; Pennings, 2012; Borucki y col., 2013; Kutnjak y col., 2017).

6.4 La adaptación a 43 °C a partir de una población ancestral más diversa produce un cuello de botella poblacional

El análisis mediante NGS de una línea representativa de cada conjunto, C43_{p2.1} y C43_{p25.3}, muestra que, cuando la adaptación a 43 °C tiene lugar a partir del ancestro más homogéneo (Q β _{p2}), en todas las regiones analizadas se producen claros cambios en la distribución de haplotipos en función de su número de orden, que en todos los casos conllevan una disminución de la frecuencia del haplotipo más representado. Sin embargo, en el caso de partir de la población ancestral Q β _{p25}, esa disminución es menor o incluso pueden producirse aumentos, como en el caso del amplicón 2, en el que el haplotipo más representado pasa de una frecuencia del 55 % en Q β _{p25} al 74 % en el pase 60 de C43_{p25.1}, lo que podría indicar una disminución de la diversidad en esta región. Las distribuciones de haplotipos en función de la distancia de Hamming también muestran que, a pase número 60, hay una mayor proporción de lecturas a distancia de Hamming 0 en el caso de la línea C43_{p25.1}, que de la C43_{p2.3}, encontrándose de nuevo las mayores diferencias en el caso del amplicón 2. En conjunto los resultados muestran que la población C43_{p25.1} está distribuida en una región del espacio de secuencias menos amplia que la ocupada por C43_{p2.3}.

En todos los casos, excepto en el amplicón 2 de C43_{p25.1}, los valores de los índices de diversidad basados en abundancia muestran un aumento a lo largo de la adaptación (Fig. 35). En el pase número 60, la línea C43_{p2.3} alcanza siempre valores mayores que la línea C43_{p25.1}, sugiriendo que, cuando existe menos diversidad genética preexistente, la estructura de la población experimenta cambios mayores a lo largo de la adaptación, probablemente debido a que la selección natural actúa sobre la diversidad generada *de novo*. Sin embargo, en una población más heterogénea es más probable que se pueda utilizar parte de esa diversidad preexistente para adaptaciones posteriores, lo cual podría incluso conducir a cuellos de botella poblacionales si existe ya algún genoma con claras ventajas sobre el resto.

El acusado descenso en la representación de las mutaciones A2187C, C2201U y C2228U, que estaban presentes en frecuencias apreciables en la población ancestral Q β _{p25} y, sin embargo, prácticamente son indetectables a pase número 10 en la línea C43_{p25.1} estaría de acuerdo con la idea de la existencia de un genoma minoritario que desplaza a otros a lo largo de la adaptación. La importancia de los genomas minoritarios en la adaptación de los virus RNA está bien establecida en numerosos estudios (Duarte y col., 1994; Cuevas y col., 2005; Schulte y Andino 2014; Bordería y col., 2015; Combe y col., 2015) y está en el origen de su elevada capacidad de adaptación (Ciota y col., 2007; 2012) y de la memoria que conservan las cuasiespecies sobre las situaciones ambientales a las que han hecho frente en el pasado (Ruíz-Jarabo y col., 2000; Briones y col., 2003; 2006).

Es notable que las mutaciones que desaparecen durante la adaptación de la línea C43_{p25.1} estén todas en la región 2, que es en la que parecía que la selección natural actuaba con más intensidad durante la evolución a 37 °C. Consistente con esta idea, ninguna de las mutaciones identificadas en esta región de Q β _{p25}, aumenta sus frecuencias a lo largo de la adaptación a 43 °C, algo que sí hacen 4 de las 5 mutaciones que estaban por encima de ese umbral en la región 1. Esto contrasta con lo observado cuando la adaptación ocurre a partir de la población ancestral Q β _{p2}, en la que las mutaciones que aumentan sus frecuencias se distribuyen en las tres regiones analizadas.

Respecto a las mutaciones que aparecen *de novo* durante la adaptación (o al menos que no estaban en frecuencias superiores al 0.5 % en la población ancestral) encontramos algunas que superan el 5 % solo de forma transitoria, otras que se mantienen en niveles no muy altos (< del 20 %) y otras que incluso llegan a fijarse. Estas mutaciones no muestran coincidencias en las dos líneas analizadas.

En C43_{p25.1} destaca el rápido ascenso de la mutación A2222C, que a pase número 10 ya está fijada, como ya se había comprobado durante el análisis de las secuencias consenso. La combinación A2222C, A1088G y U3311G, que están entre las primeras mutaciones fijadas en todas las líneas C43_{p25}, podría estar presente en algún genoma minoritario de la población ancestral que ya aporta ventajas a 43 °C. La mutación A2460C también experimenta un rápido ascenso para casi desaparecer más adelante. El hecho de que esta mutación que nunca se ha encontrado en los experimentos de adaptación de Q β a 43 °C alcance frecuencias tan altas probablemente sea debido a que está presente en algunos de los genomas que contienen A1088G, A2222C y U3311G (Tabla 24). A2460C desaparece en etapas más tardías de la evolución, indicando que un genoma con una combinación mejor de mutaciones desplaza al anterior.

En el caso de C43_{p2.3} hay que señalar el ascenso gradual de A1088G y el más brusco de U3402C. Aunque estas dos mutaciones no alcanzaron la frecuencia del 0.5 % en la población ancestral, Q β _{p2}, sí lo hicieron en Q β _{p25}, lo que sugiere que podrían estar presentes en Q β _{p2} en proporción suficiente para favorecer su rápida selección. El análisis de los datos muestra que efectivamente ambas mutaciones están en una frecuencia del 0.2 %. Es destacable que el rápido ascenso de las mutaciones U3402C y A1088G no conduce en ningún caso a descensos de los indicadores de diversidad basados en abundancia (como se veía en el caso del amplicón 2 en C43_{p2.3}), lo que indica que esta población continúa diversificándose en los pases posteriores en un grado mayor al que lo hace C43_{p25.1}. Este resultado es coherente con que la línea C43_{p2.3} a pase número 10 ha experimentado escasas ganancias en *fitness* y, por tanto, necesita continuar adquiriendo mutaciones que mejoren su capacidad replicativa a alta temperatura. Por el contrario, la línea C43_{p25.1} presenta los mayores ascensos de *fitness* entre los pases 1 y 10, por lo que no necesita experimentar tantos cambios a partir de este pase.

De acuerdo con lo anterior, una observación interesante es que en la línea C43_{p2.3} los cinco genomas más representados a pase 60 tienen frecuencias más similares que los cinco más representados a pase 10, en los que se observa una gran diferencia entre la frecuencia del más representado y los siguientes. Este hecho podría traducirse en

aumentos de complejidad en esta población entre los pases 10 y 60. Las diferencias entre las frecuencias de los haplotipos más representados en los pases 10 y 60 son mucho menos notorias en el caso de la línea C43_{P25.1}, lo que puede indicar que experimenta menos cambios en su estructura a partir del pase 10, que es cuando los valores de *fitness* también cambian menos.

Respecto a los índices de diversidad basados en incidencia, la densidad de haplotipos muestra cierta tendencia a aumentar a lo largo de la adaptación, mientras que la frecuencia de mutación mínima apenas presenta variaciones. Los valores alcanzados a pase número 60 para ambos parámetros son mayores para la línea C43_{P25.1} que para la línea C43_{P2.3}, indicando que la producción de mutaciones y su combinación en haplotipos diferentes continúa a lo largo de la adaptación de esta línea, aunque la mayoría de ellos no lleguen a hacerse mayoritarios y afecten poco a la estructura de la población.

6.5 El seguimiento de haplotipos a lo largo de la adaptación muestra la aparición de combinaciones de mutaciones que estaban presentes en genomas separados en la población ancestral

Cuando se hace un seguimiento de los cinco haplotipos más representados en cada región genómica analizada a lo largo de la adaptación de las líneas C43_{P2.3} y C43_{P25.1} se observa que, en general, el haplotipo mayoritario en las poblaciones ancestrales, que excepto en el amplicón 2 de Q β _{P25} correspondía a la secuencia sin mutaciones, es desplazado por otros que contienen uno o varios cambios genéticos. Es frecuente que mutaciones que estaban presentes en la población ancestral en haplotipos separados aparezcan reunidas en el mismo genoma, planteándose así la cuestión de si esto es debido a recombinación o a la generación recurrente de los mismos cambios. Existen ciertos trabajos previos que exploran la posibilidad de recombinación en Q β . Algunos de ellos muestran que este fago es propenso a realizar recombinación no homóloga, probablemente a través de un mecanismo que implica reacciones similares a las del procesamiento de intrones. Este tipo de recombinación resulta en la generación de RNAs que contienen fragmentos de origen viral y celular. Está claro que este mecanismo no podría explicar las diferentes combinaciones de mutaciones que observamos en nuestro análisis. La baja capacidad de la replicasa de Q β para saltar de un molde a otro hace que los eventos de recombinación sean poco comunes. (Chetverin y col., 1997). Algunas estimaciones indican que la recombinación en Q β es un millón de veces menos frecuente que en poliovirus (Palasingam y Shaklee, 1992; Chetverin y col., 2005) y puede ser compensada por la mucha mayor frecuencia de mutaciones puntuales.

Lo expresado nos hace pensar que la observada repetición de mutaciones en diferentes contextos genómicos se debe más a la generación recurrente de las mismas que a recombinación. En este sentido, algunos estudios llevados a cabo con MS2, un fago de

la misma familia que Q β aunque de un género distinto, han demostrado una alta frecuencia de recurrencia de mutaciones, lo cual podría tener como efecto la reunión en el mismo genoma de distintas mutaciones que antes estaban en genomas separados, causando así el mismo efecto que la recombinación (Bollback y Huelsenbeck, 2007; Betancourt, 2009).

Aunque los datos de las frecuencias de haplotipos en los tres amplicones analizados de cada población nos permiten intuir como se distribuyen las mutaciones en los distintos genomas y los cambios que experimentan a lo largo de la adaptación, solo la secuenciación de clones biológicos, aislados de las distintas poblaciones virales, nos permitirá comprobar la fiabilidad de nuestras suposiciones. Si, además, se realizan ensayos de *fitness* a distintas temperaturas que faciliten las correspondencias fenotipo/genotipo, podremos averiguar con mucha mayor fiabilidad si las mutaciones que desaparecen durante la adaptación (en particular A2187C, pero también C2201U o C2228U) lo hacen por tener efecto negativo en el *fitness* a 43 °C o porque en el espectro de mutantes hay un genoma con ventajas selectivas a esta temperatura que consigue desplazar a los restantes. Actualmente se están llevando a cabo experimentos en nuestro laboratorio para discernir entre estas posibilidades.

6.6 La velocidad de adaptación y las mutaciones seleccionadas a temperatura elevada dependen de la pauta de cambio

El efecto de la pauta de cambio de la temperatura en la adaptación a 43 °C se ensayó propagando la población ancestral más heterogénea, Q β _{P25}, bajo tres regímenes evolutivos que implicaban distinta velocidad de aumento de este parámetro. La adaptación fue más rápida cuanto más rápido fue el cambio de temperatura y se alcanzaron valores de *fitness* ligeramente mayores que cuando el cambio se produjo de forma más gradual. Siempre que la propagación del virus tuvo lugar a temperaturas superiores a 38 °C se produjeron aumentos de *fitness* a 43 °C, lo que sugiere que, al menos, algunas de las mutaciones que se seleccionan cuando la presión selectiva es suave también son beneficiosas cuando alcanza mayor intensidad. Cuanto más lento fue el aumento de temperatura mayor fue el número de pases necesarios para duplicar el valor de *fitness* de la población ancestral, lo que estaría de acuerdo con la selección de mutaciones de efecto pequeño cuando la presión selectiva es baja, tal y como indican algunos modelos teóricos (Bello y Waxman, 2006; Kopp y Hermisson, 2007; 2009 a, b).

Los valores de *fitness* más bajos alcanzados por las poblaciones que aumentaron su temperatura de forma progresiva también podrían ser explicados si algunas de las mutaciones adquiridas durante los pases a temperaturas más bajas tuvieran efectos negativos a temperaturas más altas o impusieran algún tipo de restricción a las nuevas mutaciones que se pueden adquirir en pases tardíos.

La determinación de las secuencias consenso a lo largo de la adaptación nuevamente muestra un comportamiento diferencial para la mutación A2187C que estaba presente

en la población ancestral $Q\beta_{P25}$ en una proporción elevada (69 %). Esta mutación permanece en todas las líneas evolucionadas de forma gradual, llegando a fijarse en varias de ellas, mientras que, como ya hemos indicado, desaparece de todas las líneas evolucionadas de forma constante a 43 °C cuyo ancestro también es la población $Q\beta_{P25}$. Estos resultados podrían interpretarse por las diferentes condiciones a las que se ve expuesta la población ancestral $Q\beta_{P25}$ en los primeros pases de su adaptación a temperatura elevada. En el caso del patrón constante (líneas $C43_{P25}$) la población se enfrenta a un cambio muy brusco en el que algunos genomas minoritarios podrían presentar ventajas y desplazar a una parte de los que existían en la población ancestral. Sin embargo, los patrones más graduales suponen la exposición de la población ancestral a temperaturas más similares a las de su evolución previa a 37 °C (40 °C en el caso de $G1_{P25}$ y 38 °C en el caso de $G2_{P25}$) que podrían no otorgar tantas ventajas a esos genomas minoritarios.

En general, en las líneas que han evolucionado de forma gradual, se observa que la mayoría de las mutaciones aumentan su representación a lo largo de la serie de pases, llegando a fijarse en algunos casos. Aunque no es habitual, algunas mutaciones que aparecen a temperaturas más bajas desaparecen a temperaturas más altas (A1930G, C2201U), lo que podría indicar que su efecto en el *fitness* depende de la temperatura, algo que, en el caso de A1930G, es coherente con los valores de su tasa de crecimiento (Fig. 48). La mutación C2201U, sin embargo, se mantuvo en otros experimentos de adaptación a temperatura elevada, lo que sugiere que ciertos contextos genómicos permiten su permanencia.

Las líneas evolucionadas de forma constante a 43 °C presentan mayor número de mutaciones en el primer evento de fijación, que además ocurre en un pase más temprano, que las líneas evolucionadas de forma gradual. A pesar de partir de la misma población ancestral, las primeras mutaciones fijadas tampoco son las mismas. Mientras que las poblaciones que han evolucionado de forma gradual inician su adaptación a partir de las mutaciones G1312A y A1930G, ninguna de las líneas evolucionadas de forma constante incluyen estas mutaciones entre las primeras que se fijan, lo que sugiere que no ofrecen grandes ventajas frente a un cambio tan drástico de temperatura como el que ocurre al pasar de 37 °C a 43 °C. Este resultado muestra claramente que la pauta de cambio ambiental es determinante en las vías adaptativas seguidas por las poblaciones.

También hay que destacar que la mayor parte de las mutaciones que se detectaron entre las primeras fijadas en las líneas $C43_{P25}$ no fueron detectadas en ninguno de los pases analizados en las líneas graduales o lo hicieron solo a pases tardíos, como es el caso de la mutación A1088G. Este resultado nuevamente sugiere que en las líneas $C43_{P25}$ la adaptación comienza a partir de genomas conteniendo combinaciones de mutaciones que ya estaban presentes en $Q\beta_{P25}$. Estos genomas probablemente desaparecen durante los primeros pases a temperaturas más bajas que tienen lugar durante la adaptación gradual, ya que en estos casos la selección actuará sobre los genomas que aporten ventajas a temperaturas moderadas. Las mutaciones fijadas en primer lugar en

las líneas C43_{P25} podrían generarse en las líneas graduales durante los pases a temperaturas más cercanas a 43 °C, pero quizás el contexto genómico generado a temperaturas inferiores no favorece su selección. El hecho de que la mayoría de las mutaciones que se fijan primero en C43_{P25} sí se detectan en las líneas C43_{P2}, que también han evolucionado de forma constante a 43 °C, aunque a partir de otra población ancestral está de acuerdo con esta explicación.

Las poblaciones que han evolucionado a través de diferentes pautas de cambio tienen diferentes historias previas y difieren en su valor de *fitness* antes de la primera exposición a 43 °C. Se observa una relación negativa entre esos valores de *fitness* y las ganancias experimentadas durante los siguientes pases a 43 °C, confirmando la observación general de que, en microorganismos, las poblaciones más aptas se adaptan más lentamente que las que lo son menos (Chou y col., 2011; Kryazhimskiy y col., 2014; Perfeito y col., 2014; Couce y Tenaillon, 2015). Este descenso en la velocidad de adaptación puede deberse a que las poblaciones mejor adaptadas ya han utilizado la mayoría de sus mutaciones beneficiosas disponibles o que estas mutaciones tienen efectos más débiles en los genotipos mejor adaptados. Esto último es una forma de epistasis conocida como epistasis de rendimientos decrecientes, la cual parece representar un tipo de interacción muy común entre mutaciones beneficiosas (Arribas y col., 2011; Khan y col., 2011). Este tipo de epistasis también podría explicar por qué algunas de las mutaciones responsables de los aumentos más grandes de *fitness* en las líneas C43_{P25} no se seleccionan cuando las poblaciones que han evolucionado con una pauta de aumento de temperatura gradual son expuestas a 43 °C.

Las líneas C43_{P25}, G1_{P25} y G2_{P25} tienen 6 mutaciones en común, que también están presentes en las poblaciones evolucionadas de forma constante a 43 °C a partir de la población ancestral Qβ_{P2}. Sin embargo, cuando solo se comparan las líneas graduales ese número aumenta a 11. Lo mismo sucede cuando se comparan las líneas que han evolucionado de forma constante (C43_{P2} y C43_{P25}), que muestran un total de 12 mutaciones en común. Estos resultados nuevamente muestran que tanto la pauta de cambio como la población ancestral son importantes a la hora de determinar las vías adaptativas y las mutaciones que se harán mayoritarias a lo largo de la adaptación.

El seguimiento mediante NGS de las frecuencias de las mutaciones que estaban presentes en la población ancestral Qβ_{P25} muestra que, a pase número 60, las dos líneas que han evolucionado de forma gradual mantienen un mayor número de mutaciones representadas en proporción mayor del 5 % de entre las que ya aparecían en la población ancestral en frecuencia mayor al 0.5 % (7 en G1_{P25.2} y 9 en G2_{P25.1}) que la línea constante (solo 4 mutaciones). El resultado está de acuerdo con el mayor número de ambientes por los que han pasado estas poblaciones hasta alcanzar los 43 °C.

El mantenimiento de más mutaciones de la población ancestral podría restringir la adquisición de nuevas mutaciones en las poblaciones que evolucionan gradualmente, explicando así la mayor convergencia entre líneas y el hecho de que sus secuencias consenso, en conjunto, presenten menor número de mutaciones que las evolucionadas de forma constante. Los valores mayores de convergencia se obtienen para la línea

G1_{P25}, lo cual está indicando que esta pauta de cambio es la más restrictiva a la hora de explorar vías adaptativas. Cuando se comparan las líneas que han evolucionado a través de distintas pautas de cambio se observan valores mayores de convergencia para los dos conjuntos de líneas graduales que cuando cada uno de ellos se compara con las líneas C43_{P25}, lo que nuevamente enfatiza la importancia del patrón de cambio en la adaptación.

6.7 La línea G1_{P25.2} presenta los valores mayores para los índices de diversidad basados en abundancia

Respecto a las variaciones de los índices de diversidad, quizás las dos observaciones más importantes son que, en el pase número 60, los basados en incidencia son menores para las dos líneas evolucionadas de forma gradual que para la evolucionada de forma constante, mientras que para los basados en abundancia, la línea G1_{P25.2} presenta siempre los valores mayores, sobre todo en el amplicón 2.

Los índices de diversidad basados en incidencia están basados en el número de entidades únicas, mutaciones o haplotipos, sin tener en cuenta su frecuencia. El hecho de que en estas poblaciones se conserven más mutaciones de las ya presentes en la población ancestral podría conducir a esos valores menores respecto a la línea evolucionada de forma constante.

Los índices de diversidad basados en abundancia, sin embargo, están muy determinados por cómo es la distribución de los haplotipos y las mutaciones teniendo en cuenta sus frecuencias. Cuando se analiza en detalle la evolución de esta población G1_{P25.2} en la región 2, se observa que el haplotipo mayoritario en la población ancestral (el que contiene A2187C) disminuye su frecuencia sin llegar a extinguirse, en favor del que contiene C2201U. Cuando la temperatura aumenta a 43 °C la frecuencia del haplotipo A2187C se sigue manteniendo alrededor del 20 %, al mismo tiempo que C2201U prácticamente desaparece y el que contiene A2187C, C2246U y C2384U llega al 40 %. Otros haplotipos que también contienen la mutación A2187C en distintos contextos genómicos se mantienen alrededor del 5 %. Esta distribución tan amplia de haplotipos hace que los índices de diversidad de la línea G1_{P25.2} tengan valores tan altos.

En las líneas C43_{P25.1} y G2_{P25.1}, sin embargo, se dan situaciones, que aunque opuestas, favorecen a un haplotipo mayoritario que tiene una frecuencia mucho más alta que el resto. En C43_{P25.1}, el haplotipo con la mutación A2187C desaparece casi totalmente en los primeros pases debido al ascenso de otro haplotipo que probablemente confiere un *fitness* mayor y que llega a ser el más representado. En G2_{P25.1}, sin embargo, la menor presión selectiva hace que el haplotipo con la mutación A2187C en ningún momento descienda sus frecuencias, por lo que continúa siendo el mayoritario en el pase número 60. En estos dos casos, tanto la variación lenta como la variación más brusca de temperatura tienen un efecto similar que es la mayor dominancia de un haplotipo sobre el resto.

6.8 La adaptación a condiciones de temperatura alternantes tiene lugar a través de las mutaciones que confieren ventajas frente a la presión selectiva más intensa

En un trabajo anterior se estudió el efecto de la alternancia de temperaturas en la evolución de Q β (Arribas y col., 2014). Los patrones que se ensayaron alternaban pases a temperaturas subóptimas (30 °C y 43 °C) con otros a la temperatura óptima del virus (37 °C). Puesto que durante los pases a 37 °C el virus podía recuperar parte de su capacidad replicativa, lo que podía tener importancia en la adaptación posterior a una nueva presión selectiva, decidimos realizar un experimento similar alternando únicamente las dos temperaturas subóptimas y durante un número mayor de pases (60 en lugar de 15). Los resultados muestran que las poblaciones alternantes consiguen aumentar su *fitness* a 43 °C, pero no a 30 °C, lo que confirma nuestro resultado previo que indicaba que la temperatura que ejerce la presión selectiva más fuerte es la que dirige la adaptación.

De acuerdo con los resultados de *fitness*, la coincidencia entre las posiciones que aparecen mutadas en las líneas alternantes y las evolucionadas de forma constante a 43 °C es mayor (12 de un total de 30) que entre las líneas alternantes y las evolucionadas de forma constante a 30 °C (2 de un total de 30). Un factor que puede influir en esta diferencia es que las líneas evolucionadas a 30 °C presentan menos posiciones mutadas que las evolucionadas de forma constante a 43 °C (11 frente a 21). También es importante señalar que de las 12 mutaciones que coinciden entre las líneas alternantes y las evolucionadas a 43 °C, 11 son no sinónimas, lo que refuerza la idea de que están implicadas en la adaptación. De hecho, entre ellas se encuentran varias (A1088G, G1312A, G1371A, A2187C, U2776C, U3311G, C3903U y G3945A) de las que están más representadas en todas las líneas evolucionadas a temperaturas más altas de la óptima.

El análisis mediante NGS de una línea representativa de cada conjunto C43_{p2}, C30_{p2} y A_{p2} (líneas 3, 1 y 2, respectivamente) muestra varios resultados interesantes. En primer lugar se observa que la línea C30_{p2}.1 presenta valores mayores de frecuencia de mutación mínima y densidad de haplotipos que la línea C43_{p2}.3. Sin embargo, los valores de los índices de diversidad basados en abundancia son mayores en la línea C43_{p2}.3. Es decir, que mientras que durante la evolución a 30 °C, hay menos mutaciones que lleguen a aparecer en la secuencia consenso, eso no parece ser debido a una menor tasa de error a temperatura más baja. Una posible explicación es que un número mayor de mutaciones tienen efectos negativos a temperatura elevada que a temperatura baja (Chen y Shakhnovich, 2010) y tienden a ser eliminadas por la selección natural. Sin embargo, probablemente también hay más mutaciones con ventajas adaptativas a 43 °C que a 30 °C, lo cual se refleja en las mutaciones detectables en las secuencias consenso. Este último resultado podría estar relacionado con las mayores dificultades para optimizar el funcionamiento de las enzimas a temperaturas bajas (Chen y Shakhnovich, 2010). Otro resultado digno de mención es que mientras que la línea alternante se asemeja mucho más a C43_{p2}.3 en cuanto a valores de *fitness* y mutaciones

que aparecen en la secuencia consenso, cuando se analizan los índices de diversidad el comportamiento varía y se asemeja más a la línea C30_{p2}.1.

Los resultados obtenidos para los valores de los índices de diversidad basados en abundancia para las líneas evolucionadas de forma constante a distintas temperaturas (30 °C, 33 °C, 37 °C, 40 °C y 43 °C) muestran una clara correlación entre la temperatura de evolución y el valor del índice considerado. Cuando se comparan estos resultados con los valores de *fitness* del virus a las distintas temperaturas, los datos parecen sugerir que cuanto mayor es la presión selectiva mayor es la complejidad de la población que se genera durante la adaptación.

No podemos asegurar que esto sea un hecho general que se cumpla para todas las presiones selectivas. Ni siquiera podemos afirmar que se cumpla si se analizan otras regiones genómicas del virus. Las presiones selectivas, tales como la presencia de drogas, para las que la adaptación es posible a través de un número pequeño de mutaciones, es menos probable que den lugar a grandes aumentos de complejidad. Sin embargo, las presiones selectivas de amplio espectro, tales como la temperatura, que afectan a prácticamente todos los procesos del virus, es más fácil que permitan la coexistencia dentro de la población de genomas conteniendo diferentes combinaciones de mutaciones que faciliten la adaptación, generándose así poblaciones más complejas, algo que hemos comprobado en un estudio reciente realizado en nuestro laboratorio (Arribas y Lázaro, enviado para publicación).

6.9 Análisis global de las mutaciones seleccionadas a temperatura elevada

Cuando se analizan las mutaciones más representadas en el conjunto de líneas evolucionadas a temperatura mayor de la óptima, encontramos un conjunto que están representadas en al menos un tercio del total. Varias de ellas [G1312A, G1371A, G3945A (en 11 líneas) U2776C (en 10 líneas) y U1295C/G (9 líneas)] se seleccionan en los tres patrones de cambio analizados y partiendo de las dos poblaciones ancestrales. Estas mutaciones, en las líneas que evolucionan gradualmente se detectan por primera vez a temperaturas entre 38 °C y 40 °C y algunas de ellas también aparecen durante la evolución constante a diversas temperaturas (U1295C/G a 37 °C, G1312A a 40 °C, G3945A a 30 °C, 33 °C, 37 °C y 40 °C, y U2776C a 40 °C). Los resultados obtenidos sugieren que esas mutaciones aportan ventajas a temperaturas moderadas, entre 38 °C y 40 °C, no siendo seleccionadas en contra a temperaturas superiores. Mención especial merece la mutación G3945A que se selecciona tanto a temperaturas por encima como por debajo de la óptima, lo cual indica que tiene un efecto beneficioso general independiente de la temperatura. Los valores de *fitness* mostrados en la Fig. 48 para las mutaciones G1312A y U2776C, muestran que la primera aporta pequeñas ventajas a 37 °C y 40 °C mientras que la segunda tiene su mayor efecto positivo a 43 °C.

Las mutaciones arriba indicadas, excepto U1295C/G, también fueron seleccionadas en el estudio de adaptación de Q β a 44 °C llevado a cabo por el grupo del Dr. Kashiwagi (Kashiwagi y col., 2014), utilizando como virus ancestral el producto de la expresión del clon infeccioso pACYCQ β (Inomata y col., 2012), que se diferencia del utilizado en este trabajo (pBRT7Q β) en la presencia en este último de las mutaciones C1257U y C2249U. Estas dos sustituciones, junto con G4A, U192C, and C2201U (que se adquirieron a temperaturas entre 37 °C y 41 °C) fueron un requerimiento necesario para que el virus obtenido a partir de la expresión de pACYCQ β se pudiera adaptar a 44 °C. La presencia en nuestro virus ancestral de C1257U y C2249U probablemente cambia el contexto genómico, permitiendo que parte de las mutaciones necesarias para la adaptación a 43 °C puedan seleccionarse a temperaturas inferiores a 40 °C.

La mutación A1088G (9 líneas), también se selecciona en todos los patrones y a partir de las dos poblaciones ancestrales. Sin embargo, a diferencia de las anteriores, esta mutación en las líneas graduales no se detecta hasta que la temperatura llega a los 43 °C. Tampoco se detecta en ninguna de las líneas evolucionadas de forma constante a temperaturas inferiores a 43 °C, sugiriendo que en esas condiciones no aporta ninguna ventaja. Estos resultados están de acuerdo con los obtenidos en el ensayo llevado a cabo con el virus resultante de la expresión del clon infeccioso pACYCQ β (Kashiwagi y col., 2014), en el que esta misma mutación no se seleccionaba hasta que la temperatura no alcanzaba los 44 °C. Los valores de *fitness* del mutante que contiene únicamente la mutación A1088G no son muy concluyentes, ya que, aunque muestra un ligero descenso a 37 °C respecto al virus sin mutaciones, a 40 °C el comportamiento de ambos virus es similar y a 43 °C solo se observa un leve aumento (Fig. 48). Una posible explicación es que el efecto positivo de la mutación A1088G sea dependiente de la presencia de otras adicionales seleccionadas en pases previos. En las líneas C43_{P25}, en las que A1088G se fija en pases tempranos, el contexto genómico favorable para su selección podría ser proporcionado por otras mutaciones que se fijan simultáneamente con ella. Un estudio posterior (Kashiwagi y col., 2018) mostró que el cambio de aminoácido producido por esta mutación (D342G en la proteína A2) reducía drásticamente la estabilidad del fago a 43.8 °C, un efecto que era en parte revertido por la presencia de mutaciones adicionales.

Mención especial merece la mutación A2187C (9 líneas) presente a pase número 60 en todas las líneas, excepto en las que evolucionan de forma constante a partir de la población Q β _{P25}, en la que, sin embargo, la mutación estaba presente como un polimorfismo detectable en la secuencia consenso. Como ya se ha indicado, una posible explicación para este resultado es que en el espectro de mutantes de la población Q β _{P25} existe algún genoma minoritario con ventajas a 43 °C, que rápidamente sustituye a los genomas con la mutación A2187C. La determinación de las secuencias consenso, junto con los resultados de NGS obtenidos para la población C43_{P25.1}, indican que esos genomas probablemente contienen las mutaciones A1088G, U2222C y U3311G. En el caso de las poblaciones que evolucionan siguiendo aumentos graduales de temperatura, que también tienen origen en Q β _{P25}, lo más probable es que esa combinación de mutaciones no ofrezca ventajas a 40 °C ni a 38 °C (o al menos que estas no superan a las

ofrecidas por los genomas que llevan la mutación A2187C), que son las temperaturas a las que se expone inicialmente la población ancestral en estos casos. La explicación es coherente con el hecho de que U2222C y U3311G tampoco se seleccionan en ninguna de las poblaciones evolucionadas de forma gradual, posiblemente porque cuando se llega a 43 °C las poblaciones ya han explorado otras vías adaptativas.

El virus generado por mutagénesis dirigida conteniendo la mutación U3311G presentó valores de *fitness* más bajos que el virus salvaje a 37 °C, 40 °C y 43 °C (Fig. 48), sugiriendo que sus ventajas a 43 °C proceden de su asociación con otras mutaciones, probablemente U2222C y A1088G según los resultados descritos arriba. Actualmente, para completar esta parte de la investigación, se están aislando clones biológicos de poblaciones en distintos momentos de su proceso evolutivo para estudiar las posibles ventajas de unas combinaciones de mutaciones frente a otras en función de la temperatura.

La mutación C3545U (7 líneas) aparece en todas las líneas evolucionadas de forma gradual, detectándose por primera vez a temperaturas entre 38 °C y 40 °C. No aparece en las constantes con origen en Q β _{P25} y solo lo hace como polimorfismo tardío en una línea con origen en Q β _{P2}. Los resultados son coherentes con los efectos ligeramente beneficiosos de esta mutación a 37 °C y perjudiciales a 43 °C (Fig. 48). El mismo comportamiento es mostrado por la mutación C1806U (6 líneas), solo presente en las líneas evolucionadas de forma gradual y en algunas de las evolucionadas de forma constante a 37 °C. Estas mutaciones también aparecen en algunas de las líneas evolucionadas de forma constante a temperaturas inferiores a 43 °C (C3545U a 30 °C y C1806U a 30 °C, 33 °C y 37 °C).

Por último, hay otras mutaciones que, aunque no estén presentes en un tercio de las líneas a pase número 60, también presentan un comportamiento interesante. Las mutaciones A1930G y C2201U se seleccionan a temperaturas entre 38 °C y 40 °C en todas las líneas que evolucionan gradualmente, pero desaparecen a temperaturas superiores, sugiriendo que esas mutaciones tienen costes en *fitness* por encima de 40 °C, algo que se confirma en el ensayo realizado con el mutante dirigido conteniendo la mutación A1930G (Fig. 48). Es destacable que ambas mutaciones también aparecen en algunas de las líneas evolucionadas de forma constante a temperaturas inferiores a 43 °C.

La mayoría de las mutaciones que hemos reseñado no parecen tener efectos específicos a una temperatura concreta y podrían contribuir a aumentar el *fitness* de Q β mediante cambios en los parámetros que definen su ciclo infectivo que son favorables en varias condiciones ambientales. En concreto, algunos ensayos muestran que el aumento de la tasa de crecimiento en las líneas de Q β adaptadas a 44 °C es debido a la mayor tasa de adsorción, el acortamiento del periodo latente y el aumento del tamaño del número de virus producidos por bacteria infectada (Kashiwagi y col., 2018).

Otra observación es que las mutaciones que se han analizado tienen efectos pequeños en el *fitness* a 43 °C cuando están solas en el genoma. Un análisis de clones biológicos

de poblaciones adaptadas a 43 °C llevado a cabo en nuestro laboratorio muestra que estas están compuestas por virus conteniendo diferentes combinaciones de mutaciones que en conjunto proporcionan valores de *fitness* bastante superiores a los que aportan cada una de las mutaciones por separado (Arribas y Lázaro; enviado para publicación).

6.10 Posibles efectos de las mutaciones seleccionadas a temperatura elevada

Aunque el análisis de los mecanismos por los que las mutaciones seleccionadas a temperatura elevada proporcionan beneficios está fuera de los objetivos de esta tesis doctoral, hemos intentado poner nuestros resultados en el contexto de otros encontrados en la literatura con objeto de intentar esclarecer en parte este tema.

En primer lugar hay que tener en cuenta que Q β es un virus de RNA y, por tanto, las mutaciones pueden tener también efecto en el *fitness* al introducir cambios en la estructura terciaria del genoma o en lugares críticos para su interacción con diferentes sustratos. Las cinco mutaciones identificadas en el ensayo de Kashiwagi y col. como necesarias para que Q β se pudiera adaptar a 44 °C son silenciosas y cuatro de ellas alteraron el patrón de pares de bases del RNA viral (Kashiwagi y col., 2014), lo que muestra la importancia de la estructura de esta molécula en el *fitness*. Las mutaciones no sinónimas, especialmente cuando tienden a conservar la naturaleza del aminoácido, también podrían estar actuando al nivel de la estructura del RNA y no de las proteínas.

Entre las regiones del genoma necesarias para la interacción de sustratos se encuentran los sitios S (nucleótidos del 1247 al 1346) y M (nucleótidos del 2545 al 2867), a los cuales se une la replicasa en su forma de holoenzima a través de la proteína S1 (Meyer y col., 1981). Los análisis de microscopía electrónica indican que, aunque ambos sitios están distantes en el genoma, pueden aproximarse a través de una interacción de larga distancia (Vollenweider y col., 1976; Barrera y col., 1993). Las mutaciones, tanto en los sitios M y S como en los nucleótidos implicados en la interacción entre ambos podrían alterar la unión de la replicasa al genoma viral causando variaciones de *fitness*.

Dos mutaciones de entre las más frecuentemente seleccionadas durante la evolución a temperatura elevada, G1312A (V417I en la proteína A2) y U2776C (V141A en la replicasa) están situadas en los sitios S y M respectivamente. Los cambios de aminoácidos que implican estas mutaciones (sustitución de una valina por una isoleucina o una alanina, respectivamente) suponen la sustitución de un aminoácido neutro y apolar por otro de las mismas características, lo que podría indicar que sus efectos se ejercen a nivel del RNA. Otra mutación frecuente es U1295C/G, localizada en el sitio S y que produce los cambios F411S/C en la proteína A2. En este caso los efectos en la estructura de la proteína podrían ser más drásticos que en los anteriores, ya que suponen la sustitución de un aminoácido aromático (fenilalanina) por otro polar con capacidad para formar puentes de H (serina o cisteína). Las propiedades similares de la serina y la cisteína explicarían la frecuente coexistencia de ambas mutaciones.

Otro sitio de interacción importante en el genoma de Q β es el correspondiente a la región del operador del gen de la replicasa. Esta proteína es un producto temprano de la expresión del genoma del fago que es requerido al inicio de la infección. Sin embargo, cuando ya hay muchas copias del RNA viral, es necesario que la replicación cese para así empaquetar los genomas en nuevas partículas virales. La regulación de este proceso es a través de la unión específica de la proteína de cubierta a una estructura en forma de horquilla al comienzo del gen de la replicasa que se conoce como el operador de la traducción (Weber, 1976). Ninguna de las mutaciones encontradas en este estudio se localizan ni en la zona del operador ni en los aminoácidos de la proteína de cubierta que interaccionan con él (Rumniers y Tars, 2014), sugiriendo que durante la adaptación a temperatura elevada no ocurren cambios en este proceso regulatorio.

La mutación A2222C, localizada en la proteína A1 y presente en varias líneas es sinónima y, por tanto, también candidata a ejercer sus efectos a nivel del RNA viral. En la proximidad de la posición 2222, en ocasiones se seleccionan las mutaciones G2223A (V293I) y C2228U (sinónima) (Arribas y Lázaro, enviado para publicación), pero siempre de forma aislada, sin combinarse entre ellas. Estos datos sugieren que las mutaciones indicadas podrían alterar la estructura del RNA de forma similar y que su presencia conjunta sería negativa para el virus. Los análisis de NGS muestran que G2223A y C2228U están presentes en proporciones apreciables en la población ancestral Q β _{P25} (3 % y 9 %, respectivamente), aunque no juntas en el mismo haplotipo. Ambas mutaciones se hacen indetectables durante la adaptación a 43 °C de la línea C43_{P25.1}, al mismo tiempo que aumenta la representación de la mutación A2222C, lo que sugiere que de las tres, esta última es la que ofrece más ventajas o está presente en un contexto genómico más favorable.

Las mutaciones A1088G (D342G), U1295(C/G) (F411S/C) y G1312A (V417I) están situadas en la proteína A2. Esta proteína está implicada tanto en la unión al pili bacteriano como en la lisis de la bacteria, para lo cual es necesario que interaccione con la proteína MurA, implicada en la síntesis del péptidoglicano. Los resultados obtenidos mediante criomicroscopía muestran que la proteína A2 está presente en la cápsida en una sola copia e interaccionando con 4 dímeros de la proteína de cubierta. Uno de los residuos implicados es el aspártico en posición 342 de la proteína A2, que interaccionaría electrostáticamente con un residuo de arginina en la proteína de cubierta (Rumnieks y Tars, 2018). La sustitución del aspártico, cargado negativamente, por glicina, que es un aminoácido neutro, debilitaría la unión de la proteína A2 con las proteínas de cubierta, lo que podría traducirse en una desestabilización de la cápsida viral.

La sustitución D342G podría tener un efecto lateral que consistiría en favorecer la interacción del virus con el pili y/o la entrada del genoma viral en la célula. En Q β , la entrada del genoma viral se realiza acompañada de la proteína A2, por lo que una interacción más débil de esta proteína con la proteína de cubierta podría favorecer su liberación. Actualmente estamos llevando a cabo ensayos de adsorción y entrada del virus para verificar esta hipótesis. Los otros dos aminoácidos mutados en la proteína A2

(posiciones 411 y 417) no están implicados en la interacción con las proteínas de cubierta ni con la región catalítica de la proteína MurA, como se desprende de los resultados de cristalización del complejo A2-MurA (Cui y col., 2017). Tampoco se ha descrito que interaccionen con el RNA (Rumnieks y Tars, 2018) por lo que es difícil sugerir una función para sus mutaciones. Una posibilidad es que estén implicados en la interacción del RNA con sustratos, como se ha señalado previamente para la posición 1312.

Entre las mutaciones altamente representadas solo encontramos una en la proteína de cubierta (G1371A que produce el cambio G9S). Quizás el grupo OH de la Serina favorezca la formación de puentes de hidrógeno dando lugar a cápsidas más estables a temperatura elevada. Respecto a las mutaciones en la proteína A1, es difícil definir su efecto, dado que se desconoce la función de esta proteína. Se ha sugerido que podría interaccionar con A2 en el proceso de reconocimiento del pili, pero hasta la fecha no hay nada publicado que lo demuestre. La mutación A2187C produce un cambio muy drástico de aminoácido (una serina por una arginina) que parece ser útil a 37 °C en las condiciones estándar de replicación del virus, como se ha demostrado en este estudio y en otros llevados a cabo en nuestro laboratorio.

Por último, ninguna de las mutaciones más frecuentes localizadas en la replicasa [U2776C (V141A), U3311G (I320M), C3903U (L517F) y G3945A (G431S)] afectan al sitio catalítico del enzima ni a sus lugares de interacción con las proteínas del huésped, el RNA que se está elongando o el nuevo nucleótido entrante (Kidmose y col., 2010; Takeshita y Tomita, 2010). Una posibilidad, que también sería aplicable al resto de mutaciones, es que su función sea estabilizar la estructura terciaria de las proteínas o el RNA. A altas temperaturas, tanto el plegamiento de las proteínas como el del RNA se ven afectados, y eso puede tener como consecuencia un peor funcionamiento de los mismos (Chen y Shakhnovich, 2010) que podría ser reparado mediante la selección de mutaciones que incrementen la estabilidad. Esta selección por estabilidad podría afectar a todas las etapas del ciclo infectivo de Q β , lo cual redundaría en aumentos de *fitness* a temperatura elevada.

7. CONCLUSIONES

1. La propagación de un clon biológico del bacteriófago Q β en las condiciones estándar de nuestro laboratorio, y a la temperatura óptima para su replicación, conduce a la generación de una población más diversa y compleja, como se demuestra por los valores de varios índices de diversidad, calculados al analizar las lecturas obtenidas tras la secuenciación masiva de tres amplicones de dos poblaciones que han experimentado diferente número de pases.
2. El análisis mediante secuenciación masiva del espectro de mutantes de la población de Q β propagada a 37 °C muestra la presencia, en frecuencias bajas, de algunas de las mutaciones que habían sido previamente identificadas como responsables de la adaptación del virus a 43 °C. Estas mutaciones, que a 37 °C se mantienen en baja proporción, podrían constituir la base sobre la que la selección natural comienza a actuar cuando hay ciertos cambios de ambiente.
3. La diversidad genética preexistente acelera la adaptación a 43 °C, aunque no se alcanzan valores mayores de fitness que cuando se parte de una población menos diversa. De acuerdo con esto, el primer evento de fijación de mutaciones también ocurre más temprano cuando la diversidad inicial es mayor.
4. La adaptación a 43 °C a partir de una población ancestral diversa produce un cuello de botella poblacional, que probablemente es debido a la rápida selección de algún genoma minoritario con ventajas a 43 °C que está presente en el espectro de mutantes de la población inicial.
5. El seguimiento de haplotipos a lo largo de la adaptación a 43 °C muestra la aparición de combinaciones de mutaciones que estaban presentes en genomas separados en la población ancestral. Este hecho posiblemente se debe a la alta tasa de error de Q β , que favorece la aparición recurrente de las mismas mutaciones en distintos contextos genómicos.
6. La adaptación a 43 °C de forma gradual produce incrementos de fitness más lentos que cuando el cambio de temperatura se realiza de forma brusca. A pesar de que hay cierta coincidencia entre las mutaciones que se seleccionan al final del proceso, las vías adaptativas no son las mismas, como se demuestra por las diferencias en las mutaciones que se fijan en primer lugar y el mayor número de mutaciones de la población ancestral que se mantienen durante la pauta de cambio gradual.
7. La propagación de Q β en condiciones que alternan la replicación a dos temperaturas subóptimas favorece la adaptación a la que genera la presión selectiva más intensa y lo hace a través de las mismas mutaciones que se seleccionan durante la adaptación constante a esa misma temperatura. Sin embargo, la estructura de la población alternante se asemeja más a la evolucionada de forma constante a la temperatura que ejerce menor presión selectiva,

mostrando así que ambas temperaturas influyen en las propiedades de la población evolucionada.

8. Cuando se analizan las mutaciones comunes en las poblaciones que han evolucionado a través de distintas pautas de cambio o a partir de diferentes poblaciones ancestrales se observa que estas son mayoritariamente no sinónimas, mientras que las mutaciones exclusivas de un tratamiento o de una línea concreta implican un mayor número de cambios sinónimos.
9. La adaptación a 43 °C ocurre a través de mutaciones que también se seleccionan a temperaturas más moderadas, entre 38 °C y 40 °C (G1312A, G1371A, U2776C), o incluso a temperaturas inferiores a 37 °C (G3945A, U1295C/G). Sin embargo, otras mutaciones solo se seleccionan cuando la temperatura alcanza los 43 °C (A1088G). Las seis mutaciones mencionadas aparecen al menos en una línea de cada conjunto evolucionado de forma gradual o constante lo que demuestra su importancia en la adaptación a 43 °C.
10. La mayoría de las mutaciones seleccionadas a temperatura elevada solo producen ligeros aumentos de fitness a 43 °C cuando están solas en el genoma. Otras no tienen ningún efecto o este es incluso negativo. Estos resultados indican que la adaptación a 43 °C tiene lugar a través del efecto combinado de varias mutaciones.
11. Dos mutaciones de entre las más frecuentemente seleccionadas durante la evolución a temperatura elevada, G1312A (V417I en la proteína A2) y U2776C (V141A en la replicasa) están situadas en los sitios S y M del genoma viral, los cuáles están implicados en la interacción con la replicasa.
12. La mutación A1088G (D342G) debilitaría la unión de la proteína A2 con las proteínas de cubierta, lo que podría traducirse en una desestabilización de la cápsida viral, algo que podría afectar a la entrada del genoma del fago en la célula.

8. REFERENCIAS

1. Acevedo, A., Brodsky, L. & Andino, R. Mutational and fitness landscapes of an RNA virus revealed through population sequencing. *Nature* **505**, 686–690 (2014).
2. Alto, B. W., Wasik, B. R., Morales, N. M. & Turner, P. E. Stochastic temperatures impede RNA virus adaptation. *Evolution* **67**, 969–979 (2013).
3. Alves, J. M. *et al.* Parallel adaptation of rabbit populations to myxoma virus. *Science* **363**, 1319–1326 (2019).
4. Amor, D. R. & Solé, R. V. Catastrophic shifts and lethal thresholds in a propagating front model of unstable tumor progression. *Phys Rev E Stat Nonlin Soft Matter Phys* **90**, 022710 (2014).
5. Andino, R. & Domingo, E. Viral quasispecies. *Virology* **479–480**, 46–51 (2015).
6. Arias, A., Lázaro, E., Escarmís, C. & Domingo, E. Molecular intermediates of fitness gain of an RNA virus: Characterization of a mutant spectrum by biological and molecular cloning. (2001) doi: 10.13039/501100010677.
7. Arribas, M., Aguirre, J., Manrubia, S. & Lázaro, E. Differences in adaptive dynamics determine the success of virus variants that propagate together. *Virus Evolution* **4**, (2018).
8. Arribas, M., Cabanillas, L., Kubota, K. & Lázaro, E. Impact of increased mutagenesis on adaptation to high temperature in bacteriophage Q β . *Virology* **497**, 163–170 (2016).
9. Arribas, M., Cabanillas, L. & Lázaro, E. Identification of mutations conferring 5-azacytidine resistance in bacteriophage Q β . *Virology* **417**, 343–352 (2011).
10. Arribas, M., Kubota, K., Cabanillas, L. & Lázaro, E. Adaptation to Fluctuating Temperatures in an RNA Virus Is Driven by the Most Stringent Selective Pressure. *PLOS ONE* **9**, e100940 (2014).
11. Axelrod, V. D., Brown, E., Priano, C. & Mills, D. R. Coliphage Q beta RNA replication: RNA catalytic for single-strand release. *Virology* **184**, 595–608 (1991).
12. Barnett, D. *pezmaster31/bamtools*. (2021).
13. Barnett, D. W., Garrison, E. K., Quinlan, A. R., Strömberg, M. P. & Marth, G. T. BamTools: a C++ API and toolkit for analyzing and managing BAM files. *Bioinformatics* **27**, 1691–1692 (2011).
14. Barrera, I., Schuppli, D., Sogo, J. M. & Weber, H. Different Mechanisms of Recognition of Bacteriophage Q β Plus and Minus Strand RNAs by Q β Replicase. *Journal of Molecular Biology* **232**, 512–521 (1993).
15. Barrett, R. D. H. & Schluter, D. Adaptation from standing genetic variation. *Trends Ecol Evol* **23**, 38–44 (2008).
16. Barrick, J. E. & Lenski, R. E. Genome dynamics during experimental evolution. *Nature Reviews Genetics* **14**, 827–839 (2013).

17. Batschelet, E., Domingo, E. & Weissmann, C. The proportion of revertant and mutant phage in a growing population, as a function of mutation and growth rate. *Gene* **1**, 27–32 (1976).
18. Beaumont, H. J. E., Gallie, J., Kost, C., Ferguson, G. C. & Rainey, P. B. Experimental evolution of bet hedging. *Nature* **462**, 90–93 (2009).
19. Bell, G. & Collins, S. Adaptation, extinction and global change. *Evolutionary Applications* **1**, 3–16 (2008).
20. Bell, G. & Gonzalez, A. Adaptation and evolutionary rescue in metapopulations experiencing environmental deterioration. *Science* **332**, 1327–1330 (2011).
21. Bello, Y. & Waxman, D. Near-periodic substitution and the genetic variance induced by environmental change. *Journal of Theoretical Biology* **239**, 152–160 (2006).
22. Bernad, A., Blanco, L., Lázaro, J. M., Martín, G. & Salas, M. A conserved 3'----5' exonuclease active site in prokaryotic and eukaryotic DNA polymerases. *Cell* **59**, 219–228 (1989).
23. Betancourt, A. J. Genomewide patterns of substitution in adaptively evolving populations of the RNA bacteriophage MS2. *Genetics* **181**, 1535–1544 (2009).
24. Biebricher, C. & Eigen, M. What Is a Quasispecies? *Current topics in microbiology and immunology* **299**, 1–31 (2006).
25. Biebricher, C. K., Eigen, M. & Gardiner, W. C. Kinetics of RNA replication: plus-minus asymmetry and double-strand formation. *Biochemistry* **23**, 3186–3194 (1984).
26. Blount, Z. D., Barrick, J. E., Davidson, C. J. & Lenski, R. E. Genomic analysis of a key innovation in an experimental *Escherichia coli* population. *Nature* **489**, 513–518 (2012).
27. Bollback, J. P. & Huelsenbeck, J. P. Clonal Interference Is Alleviated by High Mutation Rates in Large Populations. *Molecular Biology and Evolution* **24**, 1397–1406 (2007).
28. Bollback, J. P. & Huelsenbeck, J. P. Parallel Genetic Evolution Within and Between Bacteriophage Species of Varying Degrees of Divergence. *Genetics* **181**, 225–234 (2009).
29. Bons, E., Bertels, F. & Regoes, R. R. Estimating the mutational fitness effects distribution during early HIV infection. *Virus Evolution* **4**, (2018).
30. Bordería, A. V. *et al.* Group Selection and Contribution of Minority Variants during Virus Adaptation Determines Virus Fitness and Phenotype. *PLoS pathogens* **11**, e1004838 (2015).
31. Bordería, A. V. *et al.* Initial fitness recovery of HIV-1 is associated with quasispecies heterogeneity and can occur without modifications in the consensus sequence. *PLoS One* **5**, e10319 (2010).
32. Borucki, M. K. *et al.* The Role of Viral Population Diversity in Adaptation of Bovine Coronavirus to New Host Environments. *PLOS ONE* **8**, e52752 (2013).

33. Bradwell, K., Combe, M., Domingo-Calap, P. & Sanjuán, R. Correlation Between Mutation Rate and Genome Size in Riboviruses: Mutation Rate of Bacteriophage Q β . *Genetics* **195**, 243–251 (2013).
34. Briones, C., de Vicente, A., Molina-París, C. & Domingo, E. Minority memory genomes can influence the evolution of HIV-1 quasispecies in vivo. *Gene* **384**, 129–138 (2006).
35. Briones, C., Domingo, E. & Molina-París, C. Memory in retroviral quasispecies: experimental evidence and theoretical model for human immunodeficiency virus. *J Mol Biol* **331**, 213–229 (2003).
36. Brockhurst, M. A., Colegrave, N. & Rozen, D. E. Next-generation sequencing as a tool to study microbial evolution. *Molecular Ecology* **20**, 972–980 (2011).
37. Bruger, E. L. & Marx, C. J. A decade of genome sequencing has revolutionized studies of experimental evolution. *Curr Opin Microbiol* **45**, 149–155 (2018).
38. Buckling, A., Craig Maclean, R., Brockhurst, M. A. & Colegrave, N. The Beagle in a bottle. *Nature* **457**, 824–829 (2009).
39. Bull, J. J., Badgett, M. R., Springman, R. & Molineux, I. J. Genome Properties and the Limits of Adaptation in Bacteriophages. *Evolution* **58**, 692–701 (2004).
40. Bull, J. J. *et al.* Exceptional Convergent Evolution in a Virus. *Genetics* **147**, 1497–1507 (1997).
41. Burch, C. L. & Chao, L. Epistasis and Its Relationship to Canalization in the RNA Virus ϕ 6. *Genetics* **167**, 559–567 (2004).
42. Bürger, R. & Lynch, M. Evolution and Extinction in a Changing Environment: A Quantitative-Genetic Analysis. *Evolution* **49**, 151–163 (1995).
43. Burke, M. K. *et al.* Genome-wide analysis of a long-term evolution experiment with *Drosophila*. *Nature* **467**, 587–590 (2010).
44. Burke, M. K., Liti, G. & Long, A. D. Standing Genetic Variation Drives Repeatable Experimental Evolution in Outcrossing Populations of *Saccharomyces cerevisiae*. *Mol Biol Evol* **31**, 3228–3239 (2014).
45. Cabanillas, L., Arribas, M. & Lázaro, E. Evolution at increased error rate leads to the coexistence of multiple adaptive pathways in an RNA virus. *BMC Evolutionary Biology* **13**, 11 (2013).
46. Callanan, J. *et al.* RNA Phage Biology in a Metagenomic Era. *Viruses* **10**, 386 (2018).
47. Cases-González, C., Arribas, M., Domingo, E. & Lázaro, E. Beneficial effects of population bottlenecks in an RNA virus evolving at increased error rate. *J Mol Biol* **384**, 1120–1129 (2008).
48. Caspar, D. L. & Klug, A. Physical principles in the construction of regular viruses. *Cold Spring Harb Symp Quant Biol* **27**, 1–24 (1962).
49. Cervera, H. & Elena, S. F. Genetic variation in fitness within a clonal population of a plant RNA virus. *Virus Evolution* **2**, (2016).

50. Chandler, C. H. Parallel Genome-Wide Fixation of Ancestral Alleles in Partially Outcrossing Experimental Populations of *Caenorhabditis elegans*. *G3: Genes, Genomes, Genetics* **4**, 1657–1665 (2014).
51. Chao, L. Fitness of RNA virus decreased by Muller's ratchet. *Nature* **348**, 454–455 (1990).
52. Chen, P. & Shakhnovich, E. I. Thermal Adaptation of Viruses and Bacteria. *Biophys J* **98**, 1109–1118 (2010).
53. Chetverin, A. B., Chetverina, H. V., Demidenko, A. A. & Ugarov, V. I. Nonhomologous RNA recombination in a cell-free system: evidence for a transesterification mechanism guided by secondary structure. *Cell* **88**, 503–513 (1997).
54. Chetverin, A. B., Kopein, D. S., Chetverina, H. V., Demidenko, A. A. & Ugarov, V. I. Viral RNA-directed RNA Polymerases Use Diverse Mechanisms to Promote Recombination between RNA Molecules *. *Journal of Biological Chemistry* **280**, 8748–8755 (2005).
55. Chevereau, G. *et al.* Quantifying the Determinants of Evolutionary Dynamics Leading to Drug Resistance. *PLoS Biol* **13**, e1002299 (2015).
56. Chou, H.-H., Chiu, H.-C., Delaney, N. F., Segrè, D. & Marx, C. J. Diminishing returns epistasis among beneficial mutations decelerates adaptation. *Science* **332**, 1190–1192 (2011).
57. Ciota, A. T., Ehrbar, D. J., Slyke, G. A. V., Willsey, G. G. & Kramer, L. D. Cooperative interactions in the West Nile virus mutant swarm. *BMC Evolutionary Biology* **12**, undefined-undefined (2012).
58. Ciota, A. T. *et al.* Role of the mutant spectrum in adaptation and replication of West Nile virus. *J Gen Virol* **88**, 865–874 (2007).
59. Clarke, D. K. *et al.* Genetic bottlenecks and population passages cause profound fitness differences in RNA viruses. *J Virol* **67**, 222–228 (1993).
60. Collins, S. Many Possible Worlds: Expanding the Ecological Scenarios in Experimental Evolution. *Evolutionary Biology* **38**, 3–14 (2011).
61. Collins, S. & de Meaux, J. Adaptation to different rates of environmental change in *Chlamydomonas*. *Evolution* **63**, 2952–2965 (2009).
62. Combe, M., Garijo, R., Geller, R., Cuevas, J. M. & Sanjuán, R. Single-Cell Analysis of RNA Virus Infection Identifies Multiple Genetically Diverse Viral Genomes within Single Infectious Units. *Cell Host Microbe* **18**, 424–432 (2015).
63. Couce, A. & Tenaillon, O. A. The rule of declining adaptability in microbial evolution experiments. *Front. Genet.* **6**, (2015).
64. Cox, J., Schubert, A. M., Travisano, M. & Putonti, C. Adaptive evolution and inherent tolerance to extreme thermal environments. *BMC Evolutionary Biology* **10**, 75 (2010).

65. Cuevas, J. M., Duffy, S. & Sanjuán, R. Point Mutation Rate of Bacteriophage ΦX174. *Genetics* **183**, 747–749 (2009).
66. Cuevas, J. M., Moya, A. & Sanjuán, R. Following the very initial growth of biological RNA viral clones. *Journal of General Virology* **86**, 435–443 (2005).
67. Cui, Z. *et al.* Structures of Qβ virions, virus-like particles, and the Qβ–MurA complex reveal internal coat proteins and the mechanism of host lysis. *Proc Natl Acad Sci U S A* **114**, 11697–11702 (2017).
68. Desai, M. M., Fisher, D. S. & Murray, A. W. The Speed of Evolution and Maintenance of Variation in Asexual Populations. *Current Biology* **17**, 385–394 (2007).
69. Dettman, J. R. *et al.* Evolutionary insight from whole-genome sequencing of experimentally evolved microbes. *Mol Ecol* **21**, 2058–2077 (2012).
70. Domingo, E., Biebricher, C., Eigen, M. & Holland, J. J. *Quasispecies and RNA Virus Evolution: Principles and Consequences*. (Landes Bioscience, 2001).
71. Domingo, E., Escarmís, C., Menéndez-Arias, L. & Holland, J. J. Viral Quasispecies and Fitness Variations. In *Origin and Evolution of Viruses* (eds. Domingo, E., Parrish, C. R. & Holland, J. J.) (Academic Press., 1999).
72. Domingo, E. & Holland, J. J. RNA virus mutations and fitness for survival. *Annu Rev Microbiol* **51**, 151–178 (1997).
73. Domingo, E. *et al.* Viral quasispecies and the problem of vaccine-escape and drug-resistant mutants. *Prog Drug Res* **48**, 99–128 (1997).
74. Domingo, E., Sabo, D., Taniguchi, T. & Weissmann, C. Nucleotide sequence heterogeneity of an RNA phage population. *Cell* **13**, 735–744 (1978).
75. Domingo, E. Virus Evolution. In *Fields I* (eds. Knipe, D. M. & Howley, P. M.) Vol. 12 389–421 (Lippincott Williams & Wilkins., 2007).
76. Domingo, E., Dávila, M. & Ortín, J. Nucleotide sequence heterogeneity of the RNA from a natural population of foot-and-mouth-disease virus. *Gene* **11**, 333–346 (1980).
77. Domingo, E., de Ávila, A. I., Gallego, I., Sheldon, J. & Perales, C. Viral fitness: history and relevance for viral pathogenesis and antiviral interventions. *Pathogens and Disease* **77**, (2019).
78. Domingo, E. *et al.* Viral Quasispecies: Dynamics, Interactions, and Pathogenesis. *Origin and Evolution of Viruses* 87–118 (2008) doi:10.1016/B978-0-12-374153-0.00004-7.
79. Domingo, E. *et al.* Basic concepts in RNA virus evolution. *The FASEB Journal* **10**, 859–864 (1996).
80. Domingo, E. & Holland, J. J. Mutation rates and rapid evolution of RNA viruses. In *Evolutionary Biology of Viruses* (ed. Morse, S. S.) 161–184 (Raven Press, 1994).
81. Domingo, E. & Perales, C. Viral quasispecies. *PLOS Genetics* **15**, e1008271 (2019).

82. Domingo, E., Sheldon, J. & Perales, C. Viral Quasispecies Evolution. *Microbiol. Mol. Biol. Rev.* **76**, 159–216 (2012).
83. Domingo-Calap, P., Pereira-Gómez, M. & Sanjuán, R. Selection for thermostability can lead to the emergence of mutational robustness in an RNA virus. *J Evol Biol* **23**, 2453–2460 (2010).
84. Domingo-Calap, P., Cuevas, J. M. & Sanjuán, R. The Fitness Effects of Random Mutations in Single-Stranded DNA and RNA Bacteriophages. *PLOS Genetics* **5**, e1000742 (2009).
85. Drake, J. W. A constant rate of spontaneous mutation in DNA-based microbes. *Proc Natl Acad Sci U S A* **88**, 7160–7164 (1991).
86. Drake, J. W. Avoiding Dangerous Missense: Thermophiles Display Especially Low Mutation Rates. *PLOS Genetics* **5**, e1000520 (2009).
87. Drake, J. W. & Holland, J. J. Mutation rates among RNA viruses. *PNAS* **96**, 13910–13913 (1999).
88. Duarte, E. A. *et al.* Subclonal components of consensus fitness in an RNA virus clone. *Journal of Virology* **68**, 4295–4301 (1994).
89. Echols, H. & Goodman, M. F. Fidelity mechanisms in DNA replication. *Annu Rev Biochem* **60**, 477–511 (1991).
90. Eigen, M. Viral quasispecies. *Sci Am* **269**, 42–49 (1993).
91. Eigen, M. & Schuster, P. The hypercycle. A principle of natural self-organization. Part A: Emergence of the hypercycle. *Naturwissenschaften* **64**, 541–565 (1977).
92. Eigen, M. Selforganization of matter and the evolution of biological macromolecules. **58**, 465–523 (1971).
93. Elena, S. F. & Lenski, R. E. Evolution experiments with microorganisms: the dynamics and genetic bases of adaptation. *Nature Reviews Genetics* **4**, 457–469 (2003).
94. Elena, S. F. & Sanjuán, R. Virus Evolution: Insights from an Experimental Approach. *Annual Review of Ecology, Evolution, and Systematics* **38**, 27–52 (2007).
95. Elena, S. F., Solé, R. V. & Sardanyés, J. Simple genomes, complex interactions: Epistasis in RNA virus. *Chaos* **20**, 026106 (2010).
96. Escarmís, C. *et al.* Rapid selection in modified BHK-21 cells of a foot-and-mouth disease virus variant showing alterations in cell tropism. *J Virol* **72**, 10171–10179 (1998).
97. Escarmís, C., Dávila, M. & Domingo, E. Multiple molecular pathways for fitness recovery of an RNA virus debilitated by operation of Muller's ratchet. *J Mol Biol* **285**, 495–505 (1999).
98. Escarmís, C., Lázaro, E., Arias, A. & Domingo, E. Repeated bottleneck transfers can lead to non-cytocidal forms of a cytopathic virus: implications for viral extinction. *J Mol Biol* **376**, 367–379 (2008).

99. Eyre-Walker, A. & Keightley, P. D. The distribution of fitness effects of new mutations. *Nat Rev Genet* **8**, 610–618 (2007).
100. Ferrer-Orta, C. *et al.* Structure of foot-and-mouth disease virus RNA-dependent RNA polymerase and its complex with a template-primer RNA. *J Biol Chem* **279**, 47212–47221 (2004).
101. Fields, P. A. Review: Protein function at thermal extremes: balancing stability and flexibility. *Comp Biochem Physiol A Mol Integr Physiol* **129**, 417–431 (2001).
102. Fisher, R. A. The correlation between relatives on the supposition of Mendelian inheritance. *Trans. R. Soc. Edin.* **52**, 399–433 (1918).
103. Fox, E. J. & Loeb, L. A. Lethal Mutagenesis: targeting the mutator phenotype in cancer. *Semin Cancer Biol* **20**, 353–359 (2010).
104. Friedberg, E. C. *et al.* *DNA repair and mutagenesis*. (ASM Press, 2006).
105. Friedman, R., Drake, J. W. & Hughes, A. L. Genome-Wide Patterns of Nucleotide Substitution Reveal Stringent Functional Constraints on the Protein Sequences of Thermophiles. *Genetics* **167**, 1507–1512 (2004).
106. Gago, S., Elena, S. F., Flores, R. & Sanjuán, R. Extremely high mutation rate of a hammerhead viroid. *Science* **323**, 1308 (2009).
107. Gallego, I. *et al.* Broad and Dynamic Diversification of Infectious Hepatitis C Virus in a Cell Culture Environment. *Journal of Virology* **94**, (2020).
108. Gallet, R., Cooper, T. F., Elena, S. F. & Lenormand, T. Measuring selection coefficients below 10^{-3} : method, questions, and prospects. *Genetics* **190**, 175–186 (2012).
109. García-Villada, L. & Drake, J. W. Experimental selection reveals a trade-off between fecundity and lifespan in the coliphage Q β . *Open Biol* **3**, 130043 (2013).
110. Ge, L., Zhang, J., Zhou, X. & Li, H. Genetic structure and population variability of tomato yellow leaf curl China virus. *J Virol* **81**, 5902–5907 (2007).
111. Gerrish, P. J. & Lenski, R. E. The fate of competing beneficial mutations in an asexual population. *Genetica* **102**, 127 (1998).
112. Gomulkiewicz, R. & Houle, D. Demographic and genetic constraints on evolution. *Am Nat* **174**, E218–229 (2009).
113. Gonzalez, A. & Bell, G. Evolutionary rescue and adaptation to abrupt environmental change depends upon the history of stress. *Philos Trans R Soc Lond B Biol Sci* **368**, (2013).
114. Good, B. H., McDonald, M. J., Barrick, J. E., Lenski, R. E. & Desai, M. M. The dynamics of molecular evolution over 60,000 generations. *Nature* **551**, 45–50 (2017).
115. Gorter, F. A., Aarts, M. M. G., Zwaan, B. J. & de Visser, J. A. G. M. Dynamics of Adaptation in Experimental Yeast Populations Exposed to Gradual and Abrupt Change in Heavy Metal Concentration. *Am Nat* **187**, 110–119 (2016).

116. Gorter, F. A. *et al.* Genomics of Adaptation Depends on the Rate of Environmental Change in Experimental Yeast Populations. *Molecular Biology and Evolution* **34**, 2613–2626 (2017).
117. Gould, S. J. *Wonderful Life*. (Norton, 1989).
118. Gregori, J. *et al.* Viral quasispecies complexity measures. *Virology* **493**, 227–237 (2016).
119. Hao, Y.-Q., Brockhurst, M. A., Petchey, O. L. & Zhang, Q.-G. Evolutionary rescue can be impeded by temporary environmental amelioration. *Ecology Letters* **18**, 892–898 (2015).
120. Hermisson, J. & Pennings, P. S. Soft sweeps: molecular population genetics of adaptation from standing genetic variation. *Genetics* **169**, 2335–2352 (2005).
121. Hill, W. G. Rates of change in quantitative traits from fixation of new mutations. *PNAS* **79**, 142–145 (1982).
122. Hofstetter, H., Monstein, H. J. & Weissmann, C. The readthrough protein A1 is essential for the formation of viable Q beta particles. *Biochim Biophys Acta* **374**, 238–251 (1974).
123. Holder, K. K. & Bull, J. J. Profiles of adaptation in two similar viruses. *Genetics* **159**, 1393–1404 (2001).
124. Holland, J. J. Replication error, quasispecies populations, and extreme evolution rates of RNA viruses. In *Emerging Viruses* (ed. Morse, S. S.) 203–218 (Oxford Univ. Press, 1993).
125. Holland, J. J., De La Torre, J. C. & Steinhauer, D. A. RNA virus populations as quasispecies. *Curr Top Microbiol Immunol* **176**, 1–20 (1992).
126. Holland, J. *et al.* Rapid evolution of RNA genomes. *Science* **215**, 1577–1585 (1982).
127. Hug, L. A. *et al.* A new view of the tree of life. *Nature Microbiology* **1**, 1–6 (2016).
128. Inomata, T. *et al.* Quantitative comparison of the RNA bacteriophage Q β infection cycle in rich and minimal media. *Arch Virol* **157**, 2163–2169 (2012).
129. Kang, L., Aggarwal, D. D., Rashkovetsky, E., Korol, A. B. & Michalak, P. Rapid genomic changes in *Drosophila melanogaster* adapting to desiccation stress in an experimental evolution system. *BMC Genomics* **17**, 233 (2016).
130. Karnik, S. & Billeter, M. The lysis function of RNA bacteriophage Qbeta is mediated by the maturation (A2) protein. *EMBO J* **2**, 1521–1526 (1983).
131. Kashiwagi, A. *et al.* Influence of adaptive mutations, from thermal adaptation experiments, on the infection cycle of RNA bacteriophage Q β . *Arch Virol* **163**, 2655–2662 (2018).
132. Kashiwagi, A., Sugawara, R., Sano Tsushima, F., Kumagai, T. & Yomo, T. Contribution of Silent Mutations to Thermal Adaptation of RNA Bacteriophage Q β . *J Virol* **88**, 11459–11468 (2014).

133. Khan, A. I., Dinh, D. M., Schneider, D., Lenski, R. E. & Cooper, T. F. Negative epistasis between beneficial mutations in an evolving bacterial population. *Science* **332**, 1193–1196 (2011).
134. Kidmose, R. T., Vasiliev, N. N., Chetverin, A. B., Andersen, G. R. & Knudsen, C. R. Structure of the Qbeta replicase, an RNA-dependent RNA polymerase consisting of viral and host proteins. *Proc Natl Acad Sci U S A* **107**, 10884–10889 (2010).
135. Knies, J. L., Izem, R., Supler, K. L., Kingsolver, J. G. & Burch, C. L. The Genetic Basis of Thermal Reaction Norm Evolution in Lab and Natural Phage Populations. *PLOS Biology* **4**, e201 (2006).
136. Kohli, J. & Grosjean, H. Usage of the three termination codons: compilation and analysis of the known eukaryotic and prokaryotic translation termination sequences. *Mol Gen Genet* **182**, 430–439 (1981).
137. Kopp, M. & Hermisson, J. Adaptation of a quantitative trait to a moving optimum. *Genetics* **176**, 715–719 (2007).
138. Kopp, M. & Hermisson, J. The genetic basis of phenotypic adaptation I: fixation of beneficial mutations in the moving optimum model. *Genetics* **182**, 233–249 (2009).
139. Kopp, M. & Hermisson, J. The genetic basis of phenotypic adaptation II: the distribution of adaptive substitutions in the moving optimum model. *Genetics* **183**, 1453–1476 (2009).
140. Kopp, M. & Matuszewski, S. Rapid evolution of quantitative traits: theoretical perspectives. *Evol Appl* **7**, 169–191 (2014).
141. Kryazhimskiy, S., Rice, D. P., Jerison, E. R. & Desai, M. M. Microbial evolution. Global epistasis makes adaptation predictable despite sequence-level stochasticity. *Science* **344**, 1519–1522 (2014).
142. Kunkel, T. A. & Alexander, P. S. The base substitution fidelity of eucaryotic DNA polymerases. Mismatching frequencies, site preferences, insertion preferences, and base substitution by dislocation. *Journal of Biological Chemistry* **261**, 160–166 (1986).
143. Kutnjak, D., Elena, S. F. & Ravnika, M. Time-Sampled Population Sequencing Reveals the Interplay of Selection and Genetic Drift in Experimental Evolution of Potato Virus Y. *Journal of Virology* **91**, (2017).
144. Lang, G. I. & Desai, M. M. The spectrum of adaptive mutations in experimental evolution. *Genomics* **104**, 412–416 (2014).
145. Llaure, A. S. & Andino, R. Quasispecies theory and the behavior of RNA viruses. *PLOS Pathogens* **6**, e1001005 (2010).
146. Lázaro, E. RNA virus evolution at variable error rate. *Future Virology* **9**, 665–677 (2014).
147. Lázaro, E., Arribas, M., Cabanillas, L., Román, I. & Acosta, E. Evolutionary adaptation of an RNA bacteriophage to the simultaneous increase in the within-host and extracellular temperatures. *Scientific Reports* **8**, 8080 (2018).

148. Lázaro, E., Escarmís, C., Pérez-Mercader, J., Manrubia, S. C. & Domingo, E. Resistance of virus to extinction on bottleneck passages: Study of a decaying and fluctuating pattern of fitness loss. *Proc Natl Acad Sci USA* **100**, 10830–10835 (2003).
149. Lee, K. H. *et al.* First-Step Mutations for Adaptation at Elevated Temperature Increase Capsid Stability in a Virus. *PLOS ONE* **6**, e25640 (2011).
150. Lenski, R. E. Experimental evolution and the dynamics of adaptation and genome evolution in microbial populations. *The ISME Journal* **11**, 2181–2194 (2017).
151. Lenski, R. E., Rose, M. R., Simpson, S. C. & Tadler, S. C. Long-Term Experimental Evolution in *Escherichia coli*. I. Adaptation and Divergence During 2,000 Generations. *The American Naturalist* **138**, 1315–1341 (1991).
152. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv: 1303.3997 [q-bio]* (2013).
153. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
154. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**, 589–595 (2010).
155. Lindsey, H. A., Gallie, J., Taylor, S. & Kerr, B. Evolutionary rescue from extinction is contingent on a lower rate of environmental change. *Nature* **494**, 463–467 (2013).
156. Loeb, T. & Zinder, N. D. A bacteriophage containing RNA. *Proc Natl Acad Sci USA* **47**, 282–289 (1961).
157. Loewe, L. Systems in Evolutionary Systems Biology. In *Encyclopedia of Evolutionary Biology* 297–318 (Elsevier, 2016). doi:10.1016/B978-0-12-800049-6.00184-0.
158. Magoč, T. & Salzberg, S. L. FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* **27**, 2957–2963 (2011).
159. Manrubia, S. C. & Lázaro, E. Viral evolution. *Physics of Life Reviews* **3**, 65–92 (2006).
160. Martínez, M. A. *et al.* Fitness alteration of foot-and-mouth disease virus mutants: measurement of adaptability of viral quasispecies. *J Virol* **65**, 3954–3957 (1991).
161. McBride, R. C., Ogbunugafor, C. B. & Turner, P. E. Robustness promotes evolvability of thermotolerance in an RNA virus. *BMC Evol Biol* **8**, 231 (2008).
162. McGee, L. W. *et al.* Payoffs, Not Tradeoffs, in the Adaptation of a Virus to Ostensibly Conflicting Selective Pressures. *PLOS Genetics* **10**, e1004611 (2014).
163. Meyer, F., Weber, H. & Weissmann, C. Interactions of Q β replicase with Q β RNA. *Journal of Molecular Biology* **153**, 631–660 (1981).
164. Mikkelsen, J. G. & Pedersen, F. S. Genetic reassortment and patch repair by recombination in retroviruses. *J Biomed Sci* **7**, 77–99 (2000).
165. Miller, C. R., Joyce, P. & Wichman, H. A. Mutational effects and population dynamics during viral adaptation challenge current models. *Genetics* **187**, 185–202 (2011).

166. Miller, C. R., Lee, K. H., Wichman, H. A. & Ytreberg, F. M. Changing folding and binding stability in a viral coat protein: a comparison between substitutions accessible through mutation and those fixed by natural selection. *PLoS One* **9**, e112988 (2014).
167. Mills, D. R., Priano, C., Merz, P. A. & Binderow, B. D. Q beta RNA bacteriophage: mapping cis-acting elements within an RNA genome. *J Virol* **64**, 3872–3881 (1990).
168. Mills, D. R., Priano, C., DiMauro, P. & Binderow, B. D. Q beta replicase: mapping the functional domains of an RNA-dependent RNA polymerase. *J Mol Biol* **205**, 751–764 (1989).
169. Miralles, R., Gerrish, P. J., Moya, A. & Elena, S. F. Clonal interference and the evolution of RNA viruses. *Science* **285**, 1745–1747 (1999).
170. Morley, V. J., Mendiola, S. Y. & Turner, P. E. Rate of novel host invasion affects adaptability of evolving RNA virus lineages. *Proc Biol Sci* **282**, 20150801 (2015).
171. Morley, V. J. & Turner, P. E. Dynamics of molecular evolution in RNA virus populations depend on sudden versus gradual environmental change. *Evolution* **71**, 872–883 (2017).
172. Moya, A., Holmes, E. C. & González-Candelas, F. The population genetics and evolutionary epidemiology of RNA viruses. *Nat Rev Microbiol* **2**, 279–288 (2004).
173. Muller, H. J. The relation of recombination to mutational advance. *Mutat Res* **106**, 2–9 (1964).
174. Murphy, F. A. Virus taxonomy. in *Fields Virology* (eds. Fields, B. N., Knipe, D. M., Howley, P. M., Chanock, R. M. & Melnick, J. L.) 15–57 (Lippincott-Raven., 1996).
175. Napoletani, D., Signore, M. & Struppa, D. C. Cancer quasispecies and stem-like adaptive aneuploidy. *F1000Res* **2**, 268 (2013).
176. Ng, K. K.-S., Arnold, J. J. & Cameron, C. E. Structure-Function Relationships Among RNA-Dependent RNA Polymerases. *Curr Top Microbiol Immunol* **320**, 137–156 (2008).
177. Novella, I. S. *et al.* Exponential increases of RNA virus fitness during large population transmissions. *Proc Natl Acad Sci U S A* **92**, 5841–5844 (1995).
178. Novella, I. S., Elena, S. F., Moya, A., Domingo, E. & Holland, J. J. Size of genetic bottlenecks leading to virus fitness loss is determined by mean initial population fitness. *J Virol* **69**, 2869–2872 (1995).
179. Olsthoorn, R. C. L. & Van Duin, J. Leviviridae-Positive Sense RNA Viruses-Positive Sense RNA Viruses (2011)-ICTV.
https://talk.ictvonline.org/ictv-reports/ictv_9th_report/positive-sense-rna-viruses-2011/w/posrna_viruses/263/leviviridae.
180. Paepe, M. D. & Taddei, F. Viruses' Life History: Towards a Mechanistic Basis of a Trade-Off between Survival and Reproduction among Phages. *PLOS Biology* **4**, e193 (2006).

181. Palasingam, K. & Shaklee, P. N. Reversion of Q beta RNA phage mutants by homologous RNA recombination. *Journal of Virology* **66**, 2435–2442 (1992).
182. Pennings, P. S. Standing Genetic Variation and the Evolution of Drug Resistance in HIV. *PLOS Computational Biology* **8**, e1002527 (2012).
183. Perales, C. & Domingo, E. Antiviral Strategies Based on Lethal Mutagenesis and Error Threshold. *Curr Top Microbiol Immunol* **392**, 323–339 (2016).
184. Perfeito, L., Sousa, A., Bataillon, T. & Gordo, I. Rates of fitness decline and rebound suggest pervasive epistasis. *Evolution* **68**, 150–162 (2014).
185. Perron, G. G., Gonzalez, A. & Buckling, A. The rate of environmental change drives adaptation to an antibiotic sink. *J Evol Biol* **21**, 1724–1731 (2008).
186. Phillips, P. C. Epistasis: the essential role of gene interactions in the structure and evolution of genetic systems. *Nat Rev Genet* **9**, 855–867 (2008).
187. Quakkelaar, E. D. *et al.* T cell line passage can select for pre-existing neutralization-sensitive variants from the quasispecies of primary human immunodeficiency virus type-1 isolates. *Virology* **359**, 92–104 (2007).
188. Quiñones-Mateu, M. E. & Arts, E. J. Virus fitness: concept, quantification, and application to HIV population dynamics. *Curr Top Microbiol Immunol* **299**, 83–140 (2006).
189. Roff, D. *Evolutionary Quantitative Genetics*. (Springer US, 1997). doi: 10.1007/978-1-4615-4080-9.
190. Rokyta, D. R., Joyce, P., Caudle, S. B. & Wichman, H. A. An empirical test of the mutational landscape model of adaptation using a single-stranded DNA virus. *Nat Genet* **37**, 441–444 (2005).
191. Ruiz-Jarabo, C. M., Arias, A., Baranowski, E., Escarmís, C. & Domingo, E. Memory in Viral Quasispecies. *J Virol* **74**, 3543–3547 (2000).
192. Rumnieks, J. & Tars, K. Crystal Structure of the Bacteriophage Q β Coat Protein in Complex with the RNA Operator of the Replicase Gene. *Journal of Molecular Biology* **426**, 1039–1049 (2014).
193. Rūmnieks, J. & Tārs, K. Protein-RNA Interactions in the Single-Stranded RNA Bacteriophages. in *Virus Protein and Nucleoprotein Complexes* (eds. Harris, J. R. & Bhella, D.) 281–303 (Springer, 2018). doi:10.1007/978-981-10-8456-0_13.
194. Sanjuán, R. Mutational fitness effects in RNA and single-stranded DNA viruses: common patterns revealed by site-directed mutagenesis studies. *Philos Trans R Soc Lond B Biol Sci* **365**, 1975–1982 (2010).
195. Sanjuán, R., Moya, A. & Elena, S. F. The contribution of epistasis to the architecture of fitness in an RNA virus. *PNAS* **101**, 15376–15379 (2004).
196. Sanjuán, R., Nebot, M. R., Chirico, N., Mansky, L. M. & Belshaw, R. Viral mutation rates. *J Virol* **84**, 9733–9748 (2010).

197. Schenk, M. F., Szendro, I. G., Salverda, M. L. M., Krug, J. & de Visser, J. A. G. M. Patterns of Epistasis between beneficial mutations in an antibiotic resistance gene. *Mol Biol Evol* **30**, 1779–1787 (2013).
198. Schoustra, S., Hwang, S., Krug, J. & de Visser, J. A. G. M. Diminishing-returns epistasis among random beneficial mutations in a multicellular fungus. *Proc Biol Sci* **283**, (2016).
199. Schulte, M. B. & Andino, R. Single-Cell Analysis Uncovers Extensive Biological Noise in Poliovirus Replication. *Journal of Virology* **88**, 6205–6212 (2014).
200. Schuster, P. Quasispecies on Fitness Landscapes. In *Quasispecies: From Theory to Experimental Systems* (eds. Domingo, E. & Schuster, P.) 61–120 (Springer International Publishing, 2016). doi: 10.1007/82_2015_469.
201. Singhal, S. *et al.* Adaptations of an RNA virus to increasing thermal stress. *PLOS ONE* **12**, e0189602 (2017).
202. Skripkin, E. A. & Jacobson, A. B. A two-dimensional model at the nucleotide level for the central hairpin of coliphage Q beta RNA. *J Mol Biol* **233**, 245–260 (1993).
203. Sobrino, F., Dávila, M., Ortín, J. & Domingo, E. Multiple genetic variants arise in the course of replication of foot-and-mouth disease virus in cell culture. *Virology* **128**, 310–318 (1983).
204. Solé, R. V. & Deisboeck, T. S. An error catastrophe in cancer? *J Theor Biol* **228**, 47-54 (2004).
205. Strelkowa, N. & Lässig, M. Clonal Interference in the Evolution of Influenza. *Genetics* **192**, 671–682 (2012).
206. Takamatsu, H. & Iso, K. Chemical evidence for the capsomeric structure of phage q beta. *Nature* **298**, 819–824 (1982).
207. Takeshita, D. & Tomita, K. Assembly of Q β viral RNA polymerase with host translational elongation factors EF-Tu and -Ts. *PNAS* (2010) doi:10.1073/pnas.1006559107.
208. Taniguchi, T., Palmieri, M. & Weissmann, C. Q β DNA-containing hybrid plasmids giving rise to Q β phage formation in the bacterial host. *Nature* **274**, 223–228 (1978).
209. Tenaillon, O. *et al.* The molecular diversity of adaptive convergence. *Science* **335**, 457–461 (2012).
210. Toprak, E. *et al.* Evolutionary paths to antibiotic resistance under dynamically sustained drug selection. *Nat Genet* **44**, 101–105 (2011).
211. Tsukada, K. *et al.* Quantitative analysis of the bacteriophage Qbeta infection cycle. *Biochim Biophys Acta* **1790**, 65–70 (2009).
212. Vale, P. F., Choisy, M., Froissart, R., Sanjuán, R. & Gandon, S. THE DISTRIBUTION OF MUTATIONAL FITNESS EFFECTS OF PHAGE ϕ X174 ON DIFFERENT HOSTS. *Evolution* **66**, 3495–3507 (2012).

213. Valle, R. P. & Morch, M. D. Stop making sense: or Regulation at the level of termination in eukaryotic protein synthesis. *FEBS Lett* **235**, 1–15 (1988).
214. Van Duin, J. V. & Tsavera, N. Single-Stranded RNA phages. in *en The Bacteriophages* (ed. Calendar, R.) (Oxford University Press, 2006).
215. Visser, J. a. G. M. D. & Rozen, D. E. Limits to adaptation in asexual populations. *Journal of Evolutionary Biology* **18**, 779–788 (2005).
216. Vollenweider, H. J., Koller, Th., Weber, H. & Weissmann, Ch. Physical mapping of Q β replicase binding sites on Q β RNA. *Journal of Molecular Biology* **101**, 367–377 (1976).
217. Wagner, N., Atsmon-Raz, Y. & Ashkenasy, G. Theoretical Models of Generalized Quasispecies. *Curr Top Microbiol Immunol* **392**, 141–159 (2016).
218. Webb, G. F. & Blaser, M. J. Dynamics of bacterial phenotype selection in a colonized host. *PNAS* **99**, 3135–3140 (2002).
219. Weber, H. The binding site for coat protein on bacteriophage Qbeta RNA. *Biochim Biophys Acta* **418**, 175–183 (1976).
220. Webster, R. G. Chapter 14 - Antigenic Variation in Influenza Viruses. in *Origin and Evolution of Viruses* (eds. Domingo, E., Webster, R. & Holland, J.) 377–390 (Academic Press, 1999). doi:10.1016/B978-012220360-2/50015-5.
221. Weiner, A. M. & Weber, K. Natural Read-through at the UGA Termination Signal of Q β Coat Protein Cistron. *Nature New Biology* **234**, 206–209 (1971).
222. Weissmann, C. The making of a phage. *FEBS Lett* **40**, suppl:S10-18 (1974).
223. Weissmann, C., Billeter, M. A., Weber, H., Goodman, H. M. & Hindley, J. Structure and function of phage RNA: a summary of current knowledge. *Basic Life Sci* **1**, 13-28 (1973).
224. Weissmann, C., Li, J., Mahal, S. P. & Browning, S. Prions on the move. *EMBO Rep* **12**, 1109–1117 (2011).
225. Whittington, A. C. & Rokyta, D. R. Biophysical Spandrels form a Hot-Spot for Kosmotropic Mutations in Bacteriophage Thermal Adaptation. *J Mol Evol* **87**, 27-36 (2019).
226. Wichman, H. A., Badgett, M. R., Scott, L. A., Boulianne, C. M. & Bull, J. J. Different Trajectories of Parallel Evolution During Viral Adaptation. *Science* **285**, 422–424 (1999).
227. Winter, R. B. & Gold, L. Overproduction of bacteriophage Q beta maturation (A2) protein leads to cell lysis. *Cell* **33**, 877–885 (1983).
228. Wolf, J. B., Brodie, E. D. I. & Wade, M. J. *Epistasis and the Evolutionary Process*. (Oxford University Press, 2000).
229. Woodcock, D. M. *et al.* RglB facilitated cloning of highly methylated eukaryotic DNA: the human L1 transposon, plant DNA, and DNA methylated in vitro with human DNA methyltransferase. *Nucleic Acids Res* **16**, 4465–4482 (1988).

230. Wright, S. Evolution in Mendelian Populations. *Genetics* **16**, 97–159 (1931).
231. Yen, P. & Papin, J. A. History of antibiotic adaptation influences microbial evolutionary dynamics during subsequent treatment. *PLoS Biol* **15**, (2017).
232. Yubero, P., Manrubia, S. & Aguirre, J. The space of genotypes is a network of networks: implications for evolutionary and extinction dynamics. *Scientific Reports* **7**, 13813 (2017).
233. Yuste, E., Bordería, A. V., Domingo, E. & López-Galíndez, C. Few Mutations in the 5' Leader Region Mediate Fitness Recovery of Debilitated Human Immunodeficiency Type 1 Viruses. *J Virol* **79**, 5421–5427 (2005).
234. Zhao, L. & Illingworth, C. J. R. Measurements of intrahost viral diversity require an unbiased diversity metric. *Virus Evol* **5**, (2019).
235. Babraham Bioinformatics - FastQC A Quality Control tool for High Throughput Sequence Data. <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>.
236. FASTX-Toolkit. http://hannonlab.cshl.edu/fastx_toolkit/.
237. Galaxy. <https://usegalaxy.org/>.
238. Illumina | Sequencing and array-based solutions for genetic research. <https://emea.illumina.com/>.
239. Virusworld. <http://virology.wisc.edu/virusworld/viruslist.php>.

9. ANEXOS

9.1 LISTA DE ABREVIATURAS

A	Adenina
A1	Proteína readthrough de Q β
A2	Proteína de maduración de Q β
Agua DEPC	Agua tratada con dietilpirocarbonato
AMV-RT	Avian Myeloblastosis Virus Reverse Transcriptase
C	Citosina
cDNA	Ácido desoxirribonucleósido copia
DNA	Ácido desoxirribonucleósido
D.O.	Densidad óptica
dHpl	Densidad de haplotipos
D _I	Diferencias por pares
dNTP	2'-desoxinucleosido-5'-trifosfato
<i>E. coli</i>	<i>Escherichia coli</i>
EDTA	Etilen diamino tetraacetato
EF	Elongation Factor
EHF	Expand High Fidelity
F _{máx}	Frecuencia de mutación máxima
F _{mín}	Frecuencia de mutación mínima
G	Guanina
Hfr	High Frequency of Recombination
Hpl	Número de haplotipos
H _s	Entropía de Shannon
H _s H	Entropía de Shannon normalizada por el número de haplotipos
H _s N	Entropía de Shannon normalizada por el número de lecturas
Kb	Kilobase
L	Longitud de lectura
LB	Luria-Bertani
M	Sitios polimórficos
Mb	Megabase
mdi	Multiplicidad de infección
N	Lecturas representadas por encima del 0.05 %
NB	Nutrient Broth
N _f	Número de ufp iniciales
NGS	Next Generation Sequencing
N _i	Número de ufp finales
nM	Número de mutaciones
N _o	Número de lecturas tras el procesamiento bioinformático
Nt	Número de lecturas totales
ON	Overnight

pb	Pares de bases
PCR	Polymerase Chain Reaction
pi	Frecuencia de la mutación i
RdRp	RNA polimerasas dependientes de RNA
RNA	Ácido ribonucleico
rpm	Revoluciones por minuto
rRNA	Ácido ribonucleico ribosómico
RT	Retrotranscriptasa
RT-PCR	Retro transcripción y PCR
SARS-Cov-2	Severe acute respiratory syndrome coronavirus 2
SC	Secuencia consenso
+ssRNA	Ácido ribonucleico de cadena sencilla de polaridad positiva
s/n/r	Sustituciones por nucleótido y ronda de replicación
T	Timina
TAE 10X	Tris acetato EDTA 10X
T _m	Temperatura de melting
Tris	2-Amino-2-hydroxymethyl-propane-1.3-diol
Tris-HCL	Tris-clorhídrico
U	Uracilo
ufp	Unidades formadoras de placas
UV	Ultra Violeta
VFA	Virus de la Fiebre Aftosa
VGT	Virus del grabado del Tabaco
VHC	Virus de la Hepatitis C
VIH-1	Virus de la Inmunodeficiencia Humana tipo 1
Vol.	Volumen
VSV	Virus de la Estomatitis Vesicular
WT	Wild Type
π	Diversidad de nucleótidos

Códigos de aminoácidos

A	Ala	Alanina	L	Leu	Leucina
R	Arg	Arginina	K	Lys	Lisina
N	Asn	Asparagina	M	Met	Metionina
D	Asp	Aspartato	F	Phe	Fenilalanina
C	Cys	Cisteína	P	Pro	Prolina
Q	Gln	Glutamina	S	Ser	Serina
E	Glu	Ácido glutámico	T	Thr	Treonina
G	Gly	Glicina	W	Trp	Triptófano
H	His	Histidina	Y	Tyr	Tirosina
I	Ile	Isoleucina	V	Val	Valina

9.2 LISTA DE FIGURAS

Figura 1: Una vista actual del árbol de la vida, que abarca la diversidad total representada por los genomas secuenciados.

Figura 2: Esquema de una cuasiespecie.

Figura 3: Paisaje de *fitness*.

Figura 4: Esquema general de un ensayo de evolución experimental.

Figura 5: Reconstrucción de la cápside del bacteriófago Q β .

Figura 6: Mapa genómico del bacteriófago Q β .

Figura 7: Vistas al microscopio de partículas de Q β .

Figura 8: Estructura cristalina del complejo entre la subunidad β y las dos proteínas del huésped a una resolución de 2.5 Å.

Figura 9: Esquema del procedimiento experimental para la propagación del bacteriófago Q β mediante el sistema de pases seriados.

Figura 10: Esquema mostrando la evolución de la población ancestral Q β_{P25} a través de distintos patrones de cambio de temperatura.

Figura 11: Esquema mostrando la evolución de la población ancestral Q β_{P2} a través de distintos patrones de cambio de temperatura.

Figura 12: Esquema de flujo del análisis bioinformático.

Figura 13: Historia evolutiva de las poblaciones ancestrales Q β_{P2} y Q β_{P25} .

Figura 14: Mapa mostrando las regiones genómicas del bacteriófago Q β analizada por secuenciación masiva.

Figura 15: Distribución de las 100 secuencias más frecuentes en las poblaciones Q β_{P2} y Q β_{P25} para cada uno de los amplicones en estudio.

Figura 16: Esquema de las diferentes secuenciaciones de la población Q β_{P25} .

Figura 17: Frecuencias de los haplotipos representados en más del 0.05 % de las lecturas obtenidas para las poblaciones Q β_{P2} y Q β_{P25} .

Figura 18: Frecuencia de haplotipos en función de su distancia de Hamming respecto al virus Q β_{WT} para las poblaciones Q β_{P2} y Q β_{P25} .

Figura 19: Frecuencia de haplotipos en función de su distancia de Hamming respecto a la secuencia consenso para las poblaciones Q β_{P2} y Q β_{P25} .

Figura 20: Densidad de haplotipos y frecuencia de mutación mínima en las poblaciones Q β_{P2} y Q β_{P25} .

Figura 21: Valores de los índices de diversidad basados en abundancia. Poblaciones ancestrales Q β_{P2} y Q β_{P25} .

Figura 22: Diferencias entre poblaciones ($Q\beta_{P25} - Q\beta_{P2}$) para la densidad de haplotipos y la Frecuencia de mutación mínima.

Figura 23: Diferencias entre poblaciones ($Q\beta_{P25} - Q\beta_{P2}$) para los valores de los índices de diversidad basados en abundancia.

Figura 24: Esquema de obtención de las líneas evolutivas C43_{P2} y C43_{P25}.

Figura 25: Dinámica de la variación del *fitness* a 43 °C en las líneas evolutivas C43_{P2} y C43_{P25}.

Figura 26: Número total de mutaciones (fijadas y polimórficas) en las secuencias consenso de las líneas C43_{P2} y C43_{P25} en función del número de pase.

Figura 27: Mutaciones en las secuencias consenso de las líneas C43_{P2} y C43_{P25} en el pase número 60 de la adaptación a 43 °C.

Figura 28: Diagrama de Venn mostrando las mutaciones coincidentes y exclusivas de las líneas evolutivas C43_{P2} y C43_{P25}.

Figura 29: Mutaciones incluidas en el cálculo de los índices de convergencia de las líneas evolutivas C43_{P2} y C43_{P25}.

Figura 30: Frecuencias de los haplotipos representados en más del 0.05 % de las lecturas obtenidas a pase número 60 en cada uno de los amplicones analizados. a) $Q\beta_{P2}$ y C43_{P2.3} b) $Q\beta_{P25}$ y C43_{P25.1}.

Figura 31: Frecuencias de los haplotipos representados en más del 0.05 % de las lecturas obtenidas a pase número 60 en cada uno de los amplicones analizados. C43_{P2.3} y C43_{P25.1}.

Figura 32: Frecuencia de haplotipos en función de su distancia de Hamming respecto a la secuencia consenso. Poblaciones C43_{P2.3} y C43_{P25.1} obtenidas a pase número 60.

Figura 33: Frecuencia de los haplotipos más representados en las poblaciones C43_{P2.3} y C43_{P25.1} obtenidas a pase número 10 y 60.

Figura 34: Densidad de haplotipos y frecuencia de mutación mínima a lo largo de la evolución de las líneas C43_{P2.3} y C43_{P25.1}.

Figura 35: Valores de los índices de diversidad basados en abundancia a lo largo de la evolución de las líneas C43_{P2.3} y C43_{P25.1}.

Figura 36: Diferencias entre poblaciones (C43_{P25.1} - C43_{P2.3}) para los valores de los índices de diversidad basados en incidencia.

Figura 37: Diferencias entre poblaciones (C43_{P25.1} - C43_{P2.3}) para los valores de los índices de diversidad basados en abundancia.

Figura 38: Capacidad replicativa del bacteriófago $Q\beta$ a distintas temperaturas.

Figura 39: Dinámica de la variación del *fitness* a 43 °C en las líneas evolutivas C43_{P25}, G1_{P25} y G2_{P25}.

Figura 40: Número total de mutaciones (fijadas y polimórficas) en las secuencias consenso de las líneas C43_{P25}, G1_{P25} y G2_{P25} en función del número de pase.

Figura 41: Mutaciones en las secuencias consenso de las líneas G1_{P25} y G2_{P25} en el pase número 60 de la adaptación a 43 °C.

Figura 42: Diagrama de Venn mostrando las mutaciones coincidentes y exclusivas de las líneas evolutivas C43_{P25}, G1_{P25} y G2_{P25}.

Figura 43: Frecuencias de los haplotipos representados en más del 0.05 % de las lecturas obtenidas a pase número 60 en cada uno de los amplicones analizados. C43_{P25}, G1_{P25.2} y G2_{P25.1}.

Figura 44: Cambios en la distribución de frecuencias de haplotipos durante los 10 primeros pases a 43 °C en cada uno de los amplicones analizados. a) Qβ_{P25} y C43_{P25.1} pase 10, b) G1_{P25.2} pase 30 y pase 40 y c) G2_{P25.1} pase 50 y pase 60.

Figura 45: Frecuencia de haplotipos en función de su distancia de Hamming respecto a la secuencia consenso. Poblaciones C43_{P25.1}, G1_{P25.2} y G2_{P25.1} obtenidas a pase número 60.

Figura 46: Densidad de haplotipos y frecuencia de mutación mínima para el ancestro Qβ_{P25} y las líneas C43_{P25.1}, G1_{P25.2} y G2_{P25.1} a pase 60.

Figura 47: Valores de los índices de diversidad basados en abundancia para el ancestro Qβ_{P25} y las líneas C43_{P25.1}, G1_{P25.2} y G2_{P25.1} a pase 60.

Figura 48: Valores de *fitness* a diferentes temperaturas de los mutantes conteniendo las mutaciones indicadas; 37 °C, 40 °C y 43 °C.

Figura 49: Valores de *fitness* de las líneas evolutivas C30_{P2}, A_{P2} y C43_{P2} en el pase número 60.

Figura 50: Dinámica de variación del *fitness* a 43 °C en las líneas evolutivas A_{P2} y C43_{P2}.

Figura 51: Diagrama de Venn mostrando las mutaciones coincidentes y exclusivas de las líneas evolutivas C30_{P2}, A_{P2} y C43_{P2}.

Figura 52: Frecuencias de los haplotipos representados en más del 0.05 % de las lecturas obtenidas a pase número 60 en cada uno de los amplicones analizados. Poblaciones C30_{P2.1}, A_{P2.2} y C43_{P2.3}.

Figura 53: Frecuencia de haplotipos en función de su distancia de Hamming respecto a la secuencia consenso en el pase número 60 de la evolución. Poblaciones C30_{P2.1}, A_{P2.2} y C43_{P2.3}.

Figura 54: Valores de la densidad de haplotipos y frecuencia de mutación mínima en la población ancestral y en el pase número 60 de las líneas evolucionadas. Población ancestral Qβ_{P2}, C30_{P2.1}, A_{P2.2} y C43_{P2.3}.

Figura 55: Valores de los índices de diversidad basados en abundancia en la población ancestral y en el pase número 60 de las líneas evolucionadas. Población ancestral Qβ_{P2}, C30_{P2.1}, A_{P2.2} y C43_{P2.3}.

Figura 56: *Fitness* de la población ancestral Qβ_{P2} a diferentes temperaturas.

Figura 57: Valores de *fitness* de las líneas evolutivas C30_{P2}, C33_{P2}, C37_{P2}, C40_{P2} y C43_{P2} a pase número 60.

Figura 58: Frecuencias de los haplotipos representados en más del 0.05 % de las lecturas obtenidas a pase número 60 del amplicón 1. Poblaciones C30_{P2.1}, C33_{P2.2}, C37_{P2.3}, C40_{P2.1} y C43_{P2.3}.

Figura 59: Frecuencia de haplotipos en función de su distancia de Hamming respecto a la secuencia consenso obtenidas a pase número 60. Poblaciones C30_{P2.1}, C33_{P2.1}, C37_{P2.3}, C40_{P2.1} y C43_{P2.3}.

Figura 60: Densidad de haplotipos y frecuencia de mutación mínima para las líneas evolutivas C30_{P2.1}, C33_{P2.1}, C37_{P2.3}, C40_{P2.1} y C43_{P2.3} (pase 60) y su correspondiente ancestro Q β _{P2}.

Figura 61: Valores de los índices de diversidad basados en abundancia para las líneas evolutivas C30_{P2.1}, C33_{P2.1}, C37_{P2.3}, C40_{P2.1} y C43_{P2.3} (pase 60) y su correspondiente ancestro Q β _{P2}.

9.3 LISTA DE TABLAS

Tabla 1: Secuencias genéticas de las parejas de cebadores utilizados para obtener los mutantes de Q β conteniendo las mutaciones indicadas.

Tabla 2: Cebadores utilizados en las reacciones de RT-PCR.

Tabla 3: Cebadores utilizados para la determinación de las secuencias consenso.

Tabla 4: Cebadores utilizados para generar los amplicones secuenciados mediante NGS.

Tabla 5: Poblaciones ancestrales y líneas evolutivas secuenciadas mediante NGS.

Tabla 6: Número de lecturas y haplotipos obtenidos tras la secuenciación masiva de cada una de las poblaciones en estudio.

Tabla 7: Número de lecturas (*N*) y de haplotipos (*Hpl*) obtenidos en la secuenciación masiva de las poblaciones Q β _{P2} y Q β _{P25}.

Tabla 8: Número de lecturas (*N*), haplotipos (*Hpl*) y mutaciones (*nM*) obtenidos para cada amplicón en cada uno de los procesos de secuenciación de Q β _{P25}.

Tabla 9: Porcentajes de coincidencia entre haplotipos.

Tabla 10: Mutaciones representadas en proporción mayor del 0.5 % en las poblaciones Q β _{P2} y Q β _{P25}.

Tabla 11: Frecuencias (%) y mutaciones de los 5 haplotipos más representados en las poblaciones ancestrales Q β _{P2} y Q β _{P25}.

Tabla 12: Valores de los índices de diversidad basados en incidencia en las poblaciones ancestrales Q β _{P2} y Q β _{P25}.

Tabla 13: Valores de los índices de diversidad basados en abundancia en las poblaciones ancestrales $Q\beta_{P2}$ y $Q\beta_{P25}$.

Tabla 14: Valores de los índices de diversidad basados en incidencia de las poblaciones $Q\beta_{P2}$ y $Q\beta_{P25}$ para los diferentes umbrales analizados.

Tabla 15: Valores de los índices de diversidad basados en abundancia de las poblaciones $Q\beta_{P2}$ y $Q\beta_{P25}$ para los diferentes umbrales analizados.

Tabla 16: Parámetros cuantitativos que describen la dinámica de la variación del *fitness* a 43 °C en las líneas evolutivas $C43_{P2}$ y $C43_{P25}$.

Tabla 17: Mutaciones y polimorfismos en las secuencias consenso de las líneas evolutivas $C43_{P2}$ y $C43_{P25}$.

Tabla 18: Mutaciones presentes en las líneas evolutivas $C43_{P2}$ y $C43_{P25}$ en el pase número 60 de la adaptación a 43 °C.

Tabla 19: Primeras mutaciones fijadas durante la evolución a 43°C de las líneas $C43_{P2}$ y $C43_{P25}$.

Tabla 20: Valores de los índices de convergencia para las poblaciones $C43_{P2}$ y $C43_{P25}$ en el pase 60 de la adaptación a 43 °C.

Tabla 21: Trayectoria seguida por las mutaciones representadas en las poblaciones ancestrales en proporción mayor o igual al 0.5 % durante la evolución de $C43_{P2.3}$ y $C43_{P25.1}$.

Tabla 22: Mutaciones que no estaban presentes en las poblaciones ancestrales en más del 0.5 % de las lecturas y que alcanzan una representación mayor o igual al 5 % durante la evolución de las poblaciones $C43_{P2.3}$ y $C43_{P25.1}$.

Tabla 23: Secuencias de los 5 haplotipos más representados a lo largo de la adaptación a 43 °C de la línea $C43_{P2.3}$.

Tabla 24: Secuencias de los 5 haplotipos más representados a lo largo de la adaptación a 43 °C de la línea $C43_{P25.1}$.

Tabla 25: Valores de los índices de diversidad basados en incidencia de las líneas evolutivas $C43_{P2.3}$ y $C43_{P25.1}$ y sus correspondientes ancestros $Q\beta_{P2}$ y $Q\beta_{P25}$.

Tabla 26: Valores de los índices de diversidad basados en abundancia de las líneas evolutivas $C43_{P2.3}$ y $C43_{P25.1}$ y sus correspondientes ancestros $Q\beta_{P2}$ y $Q\beta_{P25}$.

Tabla 27: Valores de los índices de diversidad basados en incidencia de las líneas evolutivas $C43_{P2.3}$ y $C43_{P25.1}$ para los umbrales 0.05 % y 0.1

Tabla 28: Valores de los índices de diversidad basados en abundancia de las líneas evolutivas $C43_{P2.3}$ y $C43_{P25.1}$ para los umbrales 0.05 % y 0.1 %.

Tabla 29: Parámetros cuantitativos que describen la dinámica de la variación del *fitness* a 43 °C en las líneas evolutivas C43_{P25}, G1_{P25} y G2_{P25}.

Tabla 30: Mutaciones y polimorfismos en las secuencias consenso de las líneas evolutivas G1_{P25} y G2_{P25}.

Tabla 31: Mutaciones presentes en las líneas evolutivas C43_{P25}, G1_{P25} y G2_{P25} en el pase número 60 de la adaptación a 43 °C.

Tabla 32: Primeras mutaciones fijadas durante la evolución a 43 °C de las líneas C43_{P25}, G1_{P25} y G2_{P25}.

Tabla 33: Índices de convergencia para las poblaciones C43_{P25}, G1_{P25} y G2_{P25} en el pase 60 de la adaptación a 43 °C.

Tabla 34: Trayectoria seguida por las mutaciones representadas en la población ancestral en proporción mayor o igual al 0.5 % durante la evolución de la población G1_{P25}.2.

Tabla 35: Trayectoria seguida por las mutaciones representadas en la población ancestral en proporción mayor o igual al 0.5 % durante la evolución de la población G2_{P25}.1.

Tabla 36: Mutaciones que no estaban presentes en las poblaciones ancestrales en más del 0.5 % de las lecturas y que alcanzan una representación mayor o igual al 5 % durante la evolución de las poblaciones G1_{P25}.2 y G2_{P25}.1.

Tabla 37: Secuencias de los 5 haplotipos más representados a lo largo de la adaptación de la línea G1_{P25}.2.

Tabla 38: Secuencias de los 5 haplotipos más representados a lo largo de la adaptación de la línea G2_{P25}.1.

Tabla 39: Valores de los índices de diversidad basados en incidencia para las líneas evolutivas G1_{P25}.2, G2_{P25}.1 y su correspondiente ancestro Q β _{P25}.

Tabla 40: Valores de los índices de diversidad basados en abundancia para las líneas evolutivas G1_{P25}.2, G2_{P25}.1 y su correspondiente ancestro Q β _{P25}.

Tabla 41: Número de veces que aparece representada cada mutación en los conjuntos de líneas evolutivas G1_{P25}, G2_{P25}, C43_{P2} y C43_{P25}.

Tabla 42: Número de veces que aparece representada cada mutación en los conjuntos de líneas evolutivas G1_{P25}, G2_{P25}, C43_{P2} y C43_{P25}.

Tabla 43: Número de veces que aparece representada cada mutación en las líneas evolutivas C43_{P2} y C43_{P25}, G1_{P25} y G2_{P25}.

Tabla 44: Parámetros cuantitativos que describen la dinámica de la variación del *fitness* a 43 °C en las líneas evolutivas C43_{P2} y A_{P2}.

Tabla 45: Mutaciones y polimorfismos en las secuencias consenso de las líneas evolutivas C30_{P2}, A_{P2} y C43_{P2} en el pase número 60 de la adaptación.

Tabla 46: Mutaciones presentes en las líneas evolutivas C30_{P2}, A_{P2} y C43_{P2} en el pase número 60 de la adaptación.

Tabla 47: Frecuencias alcanzadas a pase número 60 por las mutaciones representadas en la población ancestral en proporción mayor o igual al 0.5 % en las poblaciones C30_{P2.1}, A_{P2.2} y C43_{P2.3}.

Tabla 48: Mutaciones que no estaban presentes en la población ancestral en más del 0.5 % de las lecturas y que alcanzan una representación mayor o igual al 5 % en alguna de las poblaciones evolucionadas.

Tabla 49: Valores de los índices de diversidad basados en incidencia en el pase número 60 de las líneas evolutivas C30_{P2.1}, A_{P2.2} y C43_{P2.3} y su correspondiente ancestro Q β _{P2}.

Tabla 50: Valores de los índices de diversidad basados en abundancia en el pase número 60 de las líneas evolutivas C30_{P2.1}, A_{P2.2} y C43_{P2.3} y su correspondiente ancestro Q β _{P2}.

Tabla 51: Mutaciones y polimorfismos en las secuencias consenso de las líneas evolutivas C33_{P2}, C37_{P2} y C40_{P2} en el pase número 60 de la adaptación.

Tabla 52: Mutaciones que alcanzan una representación mayor o igual al 5 % a pase número 60 de la evolución de las poblaciones C30_{P2.1}, C33_{P2.1}, C37_{P2.3}, C40_{P2.1} y C43_{P2.3}.

Tabla 53: Valores de los índices de diversidad basados en incidencia de las líneas evolutivas C30_{P2.1}, C33_{P2.1}, C37_{P2.3}, C40_{P2.1} y C43_{P2.3} obtenidas a pase número 60 y su correspondiente ancestro Q β _{P2}.

Tabla 54: Valores de los índices de diversidad basados en abundancia de las líneas evolutivas C30_{P2.1}, C33_{P2.1}, C37_{P2.3}, C40_{P2.1} y C43_{P2.3} obtenidas a pase número 60 y su correspondiente ancestro Q β _{P2}.

9.4 PUBLICACIÓN

- Somovilla, P., Manrubia, S. & Lázaro, E. Evolutionary Dynamics in the RNA Bacteriophage Q β Depends on the Pattern of Change in Selective Pressures. *Pathogens* **8**, (2019).

NOTA DE MODIFICACIÓN

Durante el proceso de análisis de los resultados de esta tesis, se detectó que el contenido de la publicación que se muestra a continuación contenía un error.

Los datos de secuenciación masiva de la población C43_{P25.1} en el pase 60 de su adaptación mostraron discrepancias entre la secuencia consenso obtenida por este método y la obtenida por secuenciación Sanger, la cual fue publicada en dicho artículo (Somovilla y col., 2019). Los resultados correctos son los que se encuentran descritos en el texto de la tesis doctoral y la secuencia consenso correcta se puede ver en la Tabla 17b.

Debido a esto, hay pequeñas discrepancias entre los resultados reales y los publicados, pero estas son mínimas y en ningún caso afectan a las conclusiones generales de la tesis y de la publicación.

Actualmente se está procediendo a realizar la corrección pertinente en la revista.

Article

Evolutionary Dynamics in the RNA Bacteriophage Q β Depends on the Pattern of Change in Selective Pressures

Pilar Somovilla ^{1,2}, Susanna Manrubia ^{2,3}  and Ester Lázaro ^{1,*}¹ Centro de Astrobiología (CSIC-INTA), 28850 Torrejón de Ardoz, Madrid, Spain; psomovilla@cnb.csic.es² Centro Nacional de Biotecnología (CSIC), 28049 Madrid, Spain; smanrubia@cnb.csic.es³ Grupo Interdisciplinar de Sistemas Complejos (GISC), Madrid, Spain

* Correspondence: lazarole@cab.inta-csic.es

Received: 22 May 2019; Accepted: 16 June 2019; Published: 18 June 2019



Abstract: The rate of change in selective pressures is one of the main factors that determines the likelihood that populations can adapt to stress conditions. Generally, the reduction in the population size that accompanies abrupt environmental changes makes it difficult to generate and select adaptive mutations. However, in systems with high genetic diversity, as happens in RNA viruses, mutations with beneficial effects under new conditions can already be present in the population, facilitating adaptation. In this work, we have propagated an RNA bacteriophage (Q β) at temperatures higher than the optimum, following different patterns of change. We have determined the fitness values and the consensus sequences of all lineages throughout the evolutionary process in order to establish correspondences between fitness variations and adaptive pathways. Our results show that populations subjected to a sudden temperature change gain fitness and fix mutations faster than those subjected to gradual changes, differing also in the particular selected mutations. The life-history of populations prior to the environmental change has great importance in the dynamics of adaptation. The conclusion is that in the bacteriophage Q β , the standing genetic diversity together with the rate of temperature change determine both the rapidity of adaptation and the followed evolutionary pathways.

Keywords: RNA viruses; bacteriophage Q β ; adaptation; pattern of change; fitness dynamics; evolutionary pathways; genetic diversity

1. Introduction

Natural populations evolve in complex environments where multiple variables act interdependently, often experiencing non-uniform patterns of change. Establishing relationships among the values of environmental variables, changes in fitness, or the presence of specific mutations is, therefore, a difficult endeavor. In addition, there are many unknowns on the relevance of evolutionary history, chance, and natural selection in shaping the final evolutionary outcome. These complications can be partly dealt with in laboratory evolution experiments carried out with multiple replicas and under conditions imposed by the experimenter, facilitating the disentanglement of correspondences between genetic and phenotypic responses through time [1–5].

Many experimental evolution studies focus on adaptive dynamics following a sudden large change in an environmental condition that is kept constant [6–8]. However, natural environments rarely change in this simple way; often, environmental changes occur gradually, sometimes experiencing fluctuations that overlap the main trend. The rate of change of selective pressures may have profound consequences for the likelihood of adaptation or extinction of a population, for the followed mutational pathways, and for the maximum degree of adaptation achieved in the process [9–16].

Evolution in gradually changing environments has been modeled as a process of adaptation to a moving optimum [17–20]. The lower the rate of change, the smaller the drop in fitness the population experiences in each of the environments to which it is exposed. This means that the selective pressure under change has a low intensity at any point of the process and adaptation will occur through the selection of beneficial mutations that provide small effects on fitness [17–20]. Since these mutations are relatively abundant [21–23], there could be a diversity of accessible pathways for adaptation to gradual changes. In contrast to this, rapid environmental changes usually produce strong fitness drops that severely decrease the population size and, thus, the associated genetic diversity [24–27], making it difficult to find the scarce, large effect mutations that would make adaptation possible, unless those mutants already exist in the population, in which case the population bottleneck might help them to fix. The situation is even harsher when a particular combination of mutations is required for optimal adaptation. Due to the generally lower number of adaptive solutions to rapid changes, evolution would be more constrained under this condition than when changes take place at a slower rate [28]. Additionally, because epistatic interactions among mutations of small effect are weaker [29,30], it has also been suggested that gradual regimes would allow for populations to reach higher fitness values.

There are many experimental studies using a variety of biological systems, such as bacteria [12,31], unicellular algae [9], yeasts [14,16,32], or viruses [13,15,33], and selective pressures, such as antibiotics [12,31], heavy metals [14,16], new hosts [13,15], salt stress [32], or changes in temperature [33], that have explored the influence of the rate of environmental change on adaptation. In good agreement with the predictions of the models, in some studies the rate of population extinction was higher under more rapid environmental changes [12,31–33] due to the strong decline in population sizes and also, in some cases, because the adapted genotypes were only accessible upon a previous sequence of environments. This kind of historical contingency can occur when certain mutations that permit survival at high intensity of the selective pressure are deleterious in the absence of other mutations that can only be selected at a lower intensity of the selective pressure [12,34]. There are also examples showing that populations evolving under gradual changes reached higher fitness values [9,13,25] and/or displayed more genetic diversity among replicate lineages than those evolving under more abrupt changes [9,12]. In contrast to this, the results obtained in other studies manifested some discrepancy with the prediction of the models, showing that phenotypes obtained under rapid environmental change were fitter [31,32,35] or contained a greater diversity of mutations than those evolving at a slower pace [13]. Finally, other studies did not find differences either in the fitness or in the genetic diversity contained in populations that could be attributed to the rate of environmental change [14,16,36]. Altogether, those results suggest that the differences in evolutionary outcomes due to the rate of change in the selective pressures depend on both the biological system—with its particular mutation and replication rates, relevance of epistatic interactions, and underlying distribution of mutational effects—and on the assayed environmental condition.

In this work, we studied how a temperature increase of fixed magnitude but variable rate influenced evolutionary outcomes in an RNA bacteriophage. Specifically, we used as a model the bacteriophage Q β , of the family *Leviviridae*, which infects *Escherichia coli* strains expressing the F pilus that acts as the virus receptor [37]. Bacteriophage Q β has an RNA genome of positive polarity (such as poliovirus, foot-and-mouth disease virus, or the coronavirus causing the severe acute respiratory syndrome) of 4217 nucleotides that encodes four proteins: the A2 protein for bacterial lysis and entry, the coat protein, the A1 protein, which is present in low amounts in the capsid and expressed through incorrect reading of the stop codon of the coat protein, and the replicase that copies the RNA genome [37]. As in other RNA viruses, replication in Q β proceeds at a high error rate [38], giving rise to highly heterogeneous populations that adapt fast. The optimal temperature for replication of this virus is 37 °C (the host temperature).

Much of the difficulty in treating diseases caused by viruses, especially RNA viruses, stems from their high evolutionary potential [39–41]. Infection of new hosts, adaptation to antivirals, or the emergence of variants that cannot be neutralized by antibodies are only a few examples of the great

adaptability of RNA viruses to changes in environmental variables. Therefore, to find effective control therapies, it is necessary to improve our knowledge of their evolutionary possibilities, for which RNA bacteriophages can be an appropriate, and easy to handle, model. In this study, the temperature for Q β replication was increased from 37 °C to 43 °C either suddenly or following two patterns of gradual change. Due to the importance of the standing genetic diversity prior to adaptation [42], we also compared the evolutionary pathways followed by the virus when populations that differed in their degree of diversification from a clonal origin were propagated at 43 °C upon a sudden exposure to this temperature.

There are previous studies concerning the adaptation of Q β to replicate at a high temperature. In some of them [43–45], the used virus variant, which could be directly propagated at 43 °C, was obtained upon expression of an infectious clone (pBRT7Q β) containing a cDNA of the virus genome cloned in pBR322 [46,47]. In another work [34] carried out with a different virus variant (obtained upon expression of the infectious clone pACYCQ β [48]), the temperature had to be gradually increased (in three steps) from 37 °C to 43.6 °C degrees to avoid virus extinction. Alignment of the virus genomes cloned in pACYCQ β and pBRT7Q β showed the presence of two silent substitutions in the latter (C1257U and C2249U). These substitutions, together with G4A, U192C, and C2201U (also silent), were a necessary requirement for the virus obtained upon expression of pACYCQ β to adapt to a high temperature [34]. Since a variant that could replicate at 43 °C was needed to analyze the adaptive process of Q β as a function of the rate of temperature change, the virus populations used in this work were all generated from the infectious clone pBRT7Q β .

The main objectives of this work were: (1) to compare the dynamics of fitness gains at 43 °C and the fitness values that were reached at the end of the transfer series, and evaluate their dependence on the rate of temperature increase; (2) to analyze the similarities of adaptive pathways among evolutionary lineages subjected to the same or a different pattern of change; (3) to identify mutations specific to a particular temperature or range of temperatures; and (4) to analyze the influence of the standing genetic diversity on the dynamics of adaptation and on the followed evolutionary pathways.

2. Results

2.1. Evolution of Bacteriophage Q β Under Different Patterns of Temperature Increase

Populations Q β -t2 and Q β -t25, obtained upon propagation of a clonal virus for two and 25 transfers at 37 °C, respectively (see Materials and Methods), were the origin for further evolution at an increased temperature, following three patterns of change (also called treatments from now on), that differed in the rate at which the maximum value (43 °C) was reached (Figure 1). The sudden pattern (S) entailed an abrupt exposure to 43 °C, after which the temperature was kept constant. The remaining two patterns involved a progressive temperature change: a gradual increase of 1 °C every 10 transfers (pattern G) or a two-step increase consisting of 30 transfers at 40 °C and 30 transfers at 43 °C (pattern TS). Population Q β -t25 was propagated through the three patterns of change, giving rise to the evolutionary lineages S-t25, TS-t25, and G-t25. Population Q β -t2 was only subjected to the sudden treatment, giving rise to the evolutionary lineages S-t2. All propagations were carried out in triplicate for a total of 60 transfers at a higher-than-optimal temperature (see Materials and Methods). Note that the number of transfers experienced at 43 °C was different for different populations: 60, 30, and 10 transfers in populations evolved through patterns S, TS, and G, respectively.

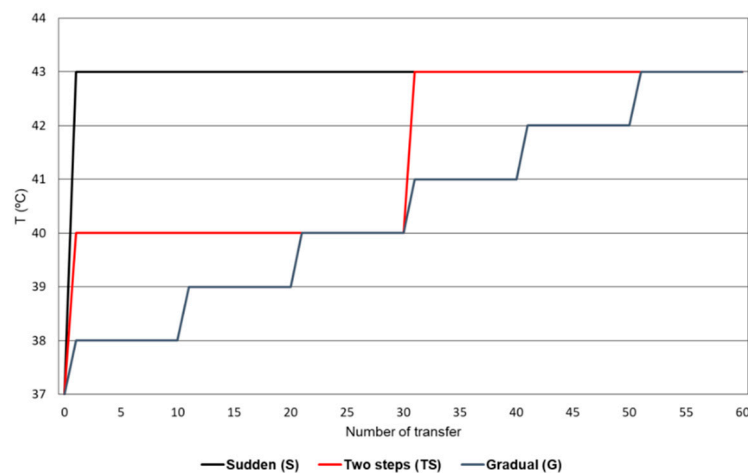


Figure 1. Patterns of temperature increase followed by the Q β evolutionary lineages. The sudden pattern (S) consisted of 60 transfers carried out at 43 °C. In the two-step pattern (TS), the virus was propagated for 30 transfers at 40 °C followed by 30 additional transfers at 43 °C. The gradual pattern (G) involved an increase of 1 °C every 10 transfers.

The intensity of the selective pressure caused by the increase of temperature was analyzed in an assay in which we determined the virus yield for population Q β -t2 at different values of this parameter (Figure 2). As expected, the highest titers were obtained at 37 °C. Temperatures above this value decreased the virus yield in an almost exponential way.

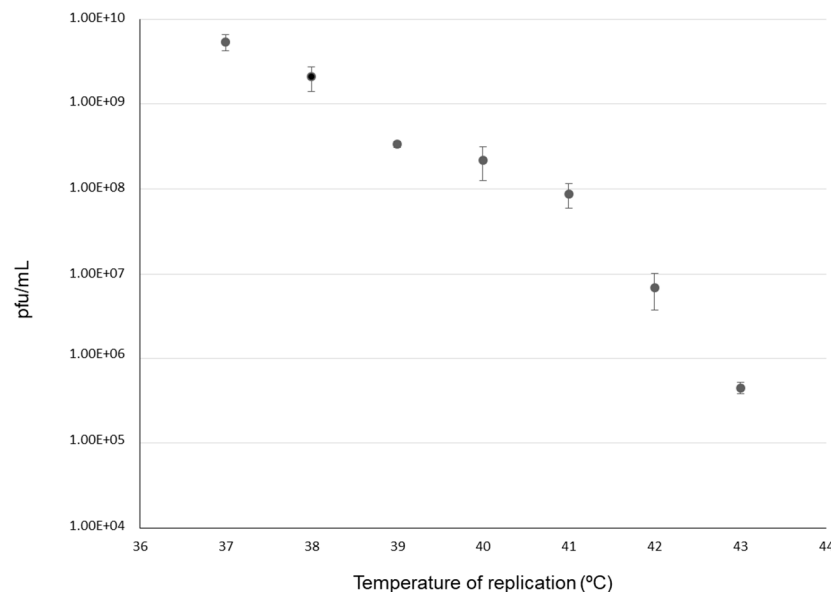


Figure 2. Sensitivity of bacteriophage Q β to increased temperature. Virus replication was assayed by infecting 1 mL *Escherichia coli* cultures (containing $\sim 10^8$ bacteria) with 10^4 plaque forming units (pfu) from the virus population Q β -t2. After incubation for 2 h at the temperatures indicated in the figure, virus supernatants were collected as described in Materials and Methods, and titrated to estimate the virus yield, which was expressed as the total number of pfu per mL of culture. Each determination corresponds to the average of two replicas and the error bars represent the standard deviation.

2.2. Dynamics of Adaptation to 43 °C

To analyze the dynamics of adaptation of Q β to increased temperature, we determined the fitness values of the virus at 43 °C in all evolutionary lineages every 10 transfers (see Materials and Methods). Values were referred to that of the corresponding ancestor population (4.7 ± 0.1 for Q β -t2 and 4.8 ± 0.7

for Q β -t25). The similarity between fitness values of populations Q β -t2 and Q β -t25 indicates that the previous evolution at 37 °C did not significantly increase fitness at 43 °C. In contrast to this, results obtained with the lineages evolved at a higher-than-optimal temperature showed that virus propagation at temperatures above 38 °C always produced fitness increases at 43 °C (Figure 3). Both the rate of temperature increase and the previous history of the ancestor population influenced fitness dynamics, as observed in the different adaptive trajectories that were followed (Figure 3). Fitness increased much faster in populations suddenly exposed to 43 °C, particularly in lineages S-t25, in which most of the increase took place during the first 10 transfers. Lineages S-t2 increased their fitness more slowly than lineages S-t25, and almost linearly until transfer number 40. In lineages evolving through the progressive treatments, the largest fitness increases took place between transfers 20 and 40 in lineages TS-t25 (the last 10 transfers at 40 °C and the first 10 transfers at 43 °C), and between transfers 10 and 20 (39 °C) and transfers 40 and 50 (42 °C) in lineages G-t25.

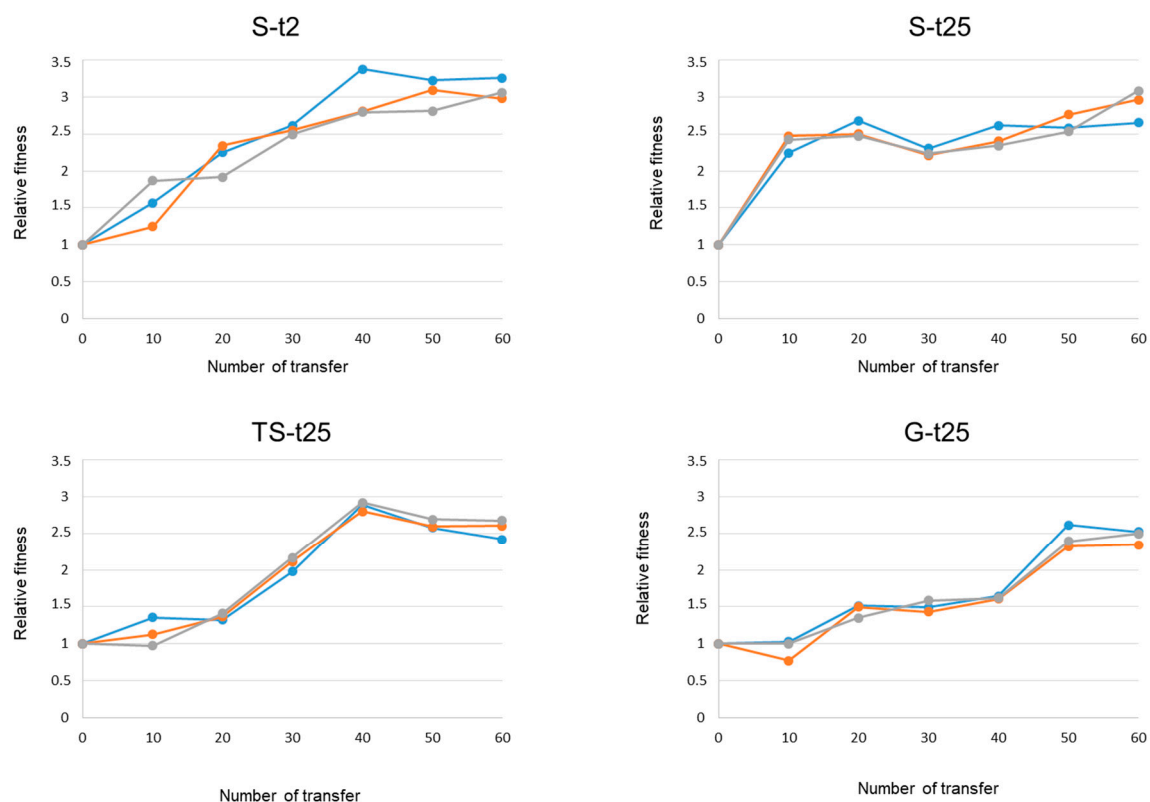


Figure 3. Fitness dynamics at 43 °C of the Q β evolutionary lineages propagated through the different patterns of temperature increase. In each treatment, the blue line represents the evolutionary lineage 1, the orange line the lineage 2, and the grey line the lineage 3. Relative fitness values were calculated as described in Materials and Methods. Each point represents the average of two replicas. The standard deviation was always lower than 10%.

To make quantitative comparisons among the different evolutionary lineages, we determined the fitness values that were reached at the end point of evolution (transfer 60), the fitness differences that were attained during the first 10 transfers at 43 °C, and the number of transfers that was necessary to duplicate the fitness of the ancestor population (that is, to reach a relative fitness value of 2). The values for all of these parameters are shown in Table 1.

Populations that were exposed directly to 43 °C (lineages S-t2 and S-t25) reached statistically similar fitness values at transfer number 60 ($p > 0.1$; Student's *t*-test). The same happened in populations that reached 43 °C progressively ($p > 0.1$ for the comparison of lineages G-t25 versus TS-t25; Student's *t*-test). On the other hand, comparison of fitness values between populations that evolved through

different patterns of temperature increase showed that sudden exposure to 43 °C allowed the virus to reach slightly higher, but significantly different, fitness values than propagation through the TS or G treatments ($p < 0.05$ for any comparison of the set of lineages S-t25 or S-t2 with either TS-t25 or G-t25; Student's *t*-test).

The variation in the fitness values during the first 10 transfers carried out at 43 °C (this is from transfer 1 to 10 in lineages S-t2 and S-t25, from transfer 30 to 40 in lineages TS-t25, and from transfer 50 to 60 in lineages G-t25) showed marked differences depending on both the ancestor population and the pattern of temperature increase (Table 1). In populations with the same ancestor (lineages S-t25, TS-t25, and G-t25), the lower the fitness value before the first exposure to 43 °C the larger the fitness gain after 10 transfers at this temperature. In populations that evolved through the sudden treatment and differed in their ancestors, fitness gains were much higher in lineages S-t25 than in S-t2 ($p < 0.05$, Student's *t*-test).

There were also important differences in the number of transfers required to duplicate the initial fitness value (Table 1). In lineages with the same origin (S-t25, TS-t25, and G-t25), the slower the temperature increase, the greater the number of transfers needed to double the initial fitness ($p < 0.001$ for all possible pair-wise comparisons between the different sets of lineages, Student's *t*-test). The importance of the ancestor population was manifested in the fact that populations S-t2 needed an average of 18.2 transfers to reach a fitness value of 2, whereas S-t25 needed only an average of 7.3 ($p < 0.01$, Student's *t*-test).

Table 1. Quantitative parameters of the fitness dynamics at 43 °C of the Q β evolutionary lineages propagated through different patterns of temperature increase.

Evolutionary Lineage	Transfer Number (Relative Fitness = 2) ¹	Relative Fitness at Transfer 60	Fitness Difference (First 10 Transfers at 43 °C)
S-t2.1	16.3	3.3	0.6
S-t2.2	16.9	3.0	0.3
S-t2.3	21.5	3.1	0.9
S-t2 (Mean \pm SD)	18.2 \pm 2.8	3.1 \pm 0.1	0.6 \pm 0.3
S-t25.1	8.1	2.7	1.2
S-t25.2	6.8	3.0	1.5
S-t25.3	7.0	3.1	1.4
S-t25 (Mean \pm SD)	7.3 \pm 0.7	2.9 \pm 0.2	1.4 \pm 0.1
TS-t25.1	30.3	2.4	0.9
TS-t25.2	28.5	2.6	0.7
TS-t25.3	27.8	2.7	0.8
TS-t25 (Mean \pm SD)	28.9 \pm 1.3	2.6 \pm 0.1	0.8 \pm 0.1
G-t25.1	43.7	2.5	-0.1
G-t25.2	45.5	2.3	0.02
G-t25.3	45.0	2.5	0.1
G-t25 (Mean \pm SD)	44.7 \pm 0.9	2.5 \pm 0.1	0.01 \pm 0.1

¹ The transfer number at which relative fitness reached a value of 2 was determined through a linear interpolation of the data shown in Figure 3.

2.3. Analysis of the Consensus Sequences of Final Q β Populations

We determined the consensus sequences of the two ancestor populations, Q β -t2 and Q β -t25, and of the 12 evolutionary lineages obtained at transfer number 60. Q β -t2 did not show any mutation relative to the Q β genome cloned in pBRT7Q β , and Q β -t25 presented only two polymorphic substitutions: A2187C (S281R in the A1 protein) and C3065U (synonymous). Both populations presented a uracil at positions 1257 and 2249, which is a characteristic of pBRT7Q β that facilitates adaptation to a

high temperature (see Introduction). In the evolved populations, we found a total of 130 mutations (including fixed and polymorphic), which were located in 35 nucleotide positions (Figure 4 and Table 2). There were no statistically significant differences in the number of mutations per genome (Figure 4) that could be attributed to the pattern of temperature increase ($p > 0.05$ for all possible pair-wise comparisons between different sets of lineages, Student's *t*-test). Mutation C3065U, which was present in the ancestor Q β -t25, was lost in all lineages founded by this population. However, since the mutation was polymorphic in Q β -t25, we cannot know whether the loss was due to a reversion or to negative selection, though we believe the latter is more plausible. Substitution A2187C, also in Q β -t25, presented a non-uniform behavior. It was lost in some lineages, whereas in others was fixed or remained polymorphic.

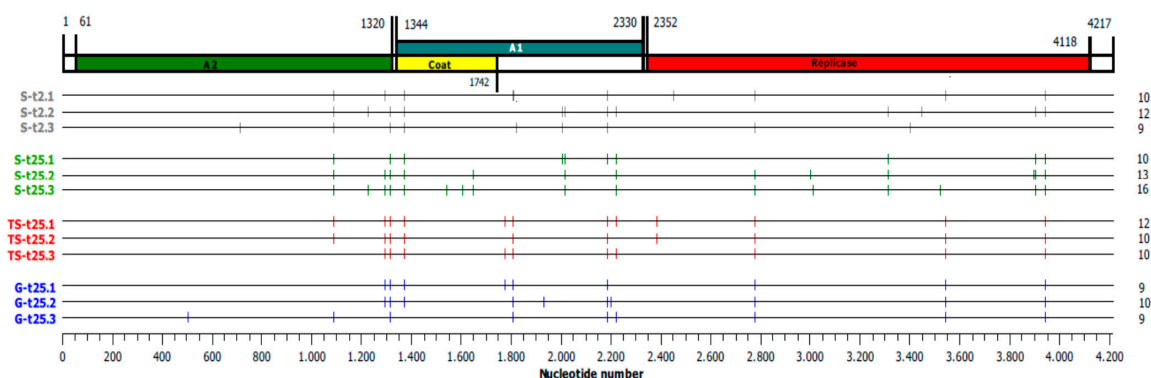


Figure 4. Mutated positions present at transfer number 60 in the consensus sequences of the Q β evolutionary lineages propagated through the different patterns of temperature increase. All mutations (fixed and polymorphic) relative to the sequence of the Q β cDNA cloned in pBRT7Q β (see Materials and Methods) are represented as vertical lines on the horizontal lines corresponding to the Q β genome. The total number of mutations in each genome is shown on the right part of the figure.

From the set of 130 mutations, 105 (located in 23 nucleotide positions) were non synonymous, which is a clear excess over the number of synonymous mutations (Table 2). Sixteen substitutions were exclusive of only one evolutionary lineage, whereas others were represented in at least half of the lineages (A1088G, U1295(C/G), G1312A, G1371A, C1806U, U2776C, C3545U, and G3945A). All but one (C3545U) of these highly represented substitutions were non synonymous and distributed similarly across all treatments (again with the exception of C3545U, which was more frequent in the progressive treatments).

The distribution of mutated positions common to and exclusive of the different treatments is represented in Figure 5. When comparing the lineages with the same origin and evolving through different patterns of temperature increase (Figure 5A), we can observe that there are some substitutions (A1088G, U1295(C/G), G1312A, G1371A, A2187C, U2776C, and G3945A) that were present in at least one lineage of each of the three sets S-t25, TS-t25, and G-t25. The same substitutions, together with six additional ones, were also represented in at least one lineage of the two sets St-2 and St-25 (Figure 5B). There were also some substitutions common to S-t2 and S-t25 and absent from the progressive treatments, and conversely (in red in Figure 5). The number of different substitutions exclusive of each set of replicate lineages was 7 for S-t25 and S-t2, 1 for TS-t25, and 3 for G-t25 (in green in Figure 5). Most of them were present in a single lineage.

Table 2. Mutations present at transfer number 60 in the sets of lineages propagated through different patterns of temperature increase.

Mutation ¹	Gene ²	Number of Lineages with the Mutation Per Treatment ³			
		S-t2	S-t25	TS-t25	G-t25
G502C (G147R)	A2				0/1
C714U (S)	A2	0/1			
A1088G (D342G)	A2	2/1	3/0	0/2	0/1
C1228U (L389F)	A2	0/1	1/0		
U1295(C/G) (F411S/C)	A2	1/0	2/0	1/2	1/1
G1312A (V417I)	A2	2/0	3/0	3/0	3/0
G1371A (G9S)	C	3/0	3/0	0/3	0/2
U1544C (S)	C		1/0		
G1607U (Q87H)	C		1/0		
C1649U (S)	C		2/0		
A1778G (S)	A1			1/1	1/0
C1806U (P154S)	A1			3/0	2/1
C1807U (P154L)	A1	1/0			
U1810C (I155T)	A1	1/0			
G1820A (S)	A1	1/0			
A1930G (Q195R)	A1				0/1
U2006G (S220R)	A1	0/2	0/1		
U2016C (F224L)	A1	1/0	2/1		
A2187C (S281R)	A1	0/3	0/1	3/0	2/1
C2201U (S)	A1				1/0
A2222C (S)	A1	1/0	3/0		
G2223A (V293I)	A1			1/1	0/1
C2384U (S)	R			0/2	
C2452U (A33V)	R	1/0			
U2776C (V141A)	R	2/0	1/1	3/0	3/0
G3002A (S)	R		1/0		
U3011C (S)	R		1/0		
U3311G (I320M)	R	1/0	3/0		
U3402C (S350P)	R	1/0			
U3447C (C365R)	R	0/1			
G3521U (E389D)	R		1/0		
C3545U (S)	R	0/1		3/0	2/1
G3896A (S)	R		0/1		
C3903U (L517F)	R	0/1	2/1		
G3945A (G531S)	R	2/0	3/0	3/0	1/2

¹ The change of amino acid is shown in brackets. S indicates synonymous mutations. Position 1295 could mutate to C or G. ² Location of the mutation in the different genes of the Q β genome. C means coat and R replicase. ³ The first number of each pair indicates the number of lineages in a given set in which the mutation is fixed. The second number refers to the number of lineages in which the same mutation is present as a polymorphism. Empty cells indicate the absence of the mutation in that particular set of lineages.

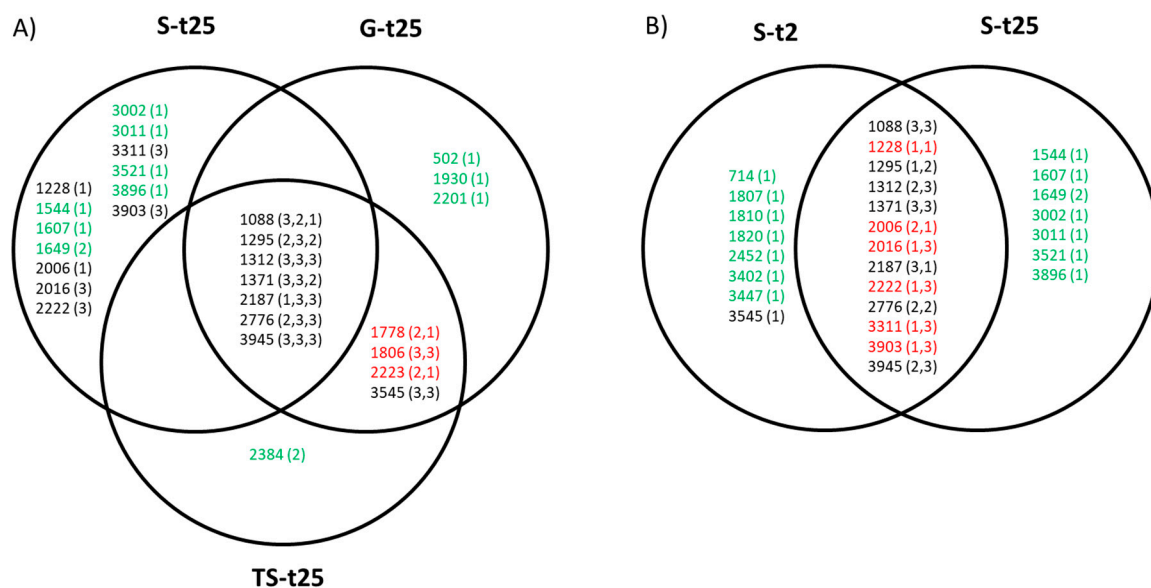


Figure 5. Representation of mutated positions at transfer number 60 in the sets of lineages propagated through different patterns of temperature increase. **(A)** Lineages with the same origin and subjected to different treatments. **(B)** Lineages subjected to the sudden treatment and differing in their ancestor population. Each circle encloses the nucleotide positions that are mutated at least in one evolutionary lineage of each set. The numbers between brackets indicate the number of times that a particular position is mutated per set of lineages evolving under the same condition. The order of the numbers for positions common to several sets is: (S-t25, TS-t25, G-t25) or (TS-t25, G-t25) in part A), and (S-t2, S-t25) in part B). Mutated positions common to S-t2 and S-t25, and non-mutated in the progressive treatments, and conversely are shown in red. Mutated positions exclusive of each set of replicate lineages are shown in green.

2.4. Evolutionary Convergence among Lineages and Treatments

The nucleotide diversity among replicate lineages was evaluated as the number of pair-wise differences between genomes divided by the number of analyzed nucleotides. The results showed that lineages S-t2 had higher nucleotide diversity than lineages S-t25 (3.2×10^{-3} vs. 2.1×10^{-3}). Evolution through any of the progressive treatments led to lower nucleotide diversity (1.4×10^{-3} for lineages G-t25 and 7×10^{-4} for lineages TS-t25) than evolution from the same ancestor and through the sudden treatment (lineages S-t25). The conclusion is that both the characteristics of the ancestor population and the pattern of temperature increase influence the nucleotide diversity among replicate lineages.

Convergence between different sets of replicate lineages was evaluated by counting the number of mutations that were present in at least one lineage of each set, and dividing it by the total number of different mutations in the two sets. Convergence between lineages evolved through a similar treatment had a value of 0.46 for lineages S-t2 versus S-t25 (sudden treatment), and of 0.73 for TS-t25 versus G-t25 (progressive treatments). That value decreased when populations evolved under any of the progressive patterns were compared to those subjected to the sudden treatment (0.26 for lineages G-t25 vs. S-t25 and 0.28 for lineages TS-t25 vs. S-t25). Altogether, the results show that the pattern of temperature change is more important than the ancestor population to determine the mutations that can be selected throughout the adaptive process and that contribute to define the degree of convergence between different treatments.

2.5. Evolution of the Consensus Sequences through Adaptation

With the aim of analyzing the dynamics of adaptation at the genetic level, we determined the consensus sequence of all evolutionary lineages every 10 transfers (Supplementary Figure S1). Representation of the total number of mutations (fixed and polymorphic) as a function of the number of

transfers (Supplementary Figure S2) allowed us to investigate whether the pace of mutation acquisition mimicked the dynamics of adaptation shown in Figure 3. The results showed that, during the first 10 transfers, lineages S-t25 acquired mutations much more rapidly than the rest ($p < 0.001$ for all comparisons between the set of lineages S-t25 with any other set; Student's *t*-test), which agrees with their faster adaptation at early stages. At transfer number 20, the differences were lower but still statistically significant ($p < 0.05$ for all comparisons between the set of lineages S-t25 with any other set; Student's *t*-test). From transfer 40 onwards, there were no significant differences in the number of mutations between any set of evolutionary lineages.

We also determined the mutations that first fixed in each evolutionary lineage and the number of transfers at which they fixed (Table 3). In populations TS-t25 and G-t25, the first mutation fixed was G1312A (in all evolutionary lineages but G-t25.2, where A1930G was the first mutation fixed), and only in one case (TS-t25.2) two mutations fixed simultaneously (A1930G, G1312A). All these substitutions fixed at transfer number 20, with the exception of G1312A in TS-t25.3 that fixed at transfer number 30. Populations evolved through the sudden treatment yielded different quantitative results. The first event of fixation of mutations involved a wider set of mutations and, in the case of lineages S-t25, it also took place earlier (at transfer number 10). Remarkably, substitutions A1930G and G1312A, which were the first to be fixed in the progressive treatments, were not found amongst the first fixed in the sudden treatment, with the only exception of G1312A in S-t2.3. Substitution A1930G was not present at any transfer in any of the lineages evolved constantly at 43 °C, and was lost or became less conspicuous in all lineages evolving through the progressive treatments. Remarkably, most of the substitutions among those first fixed in lineages S-t25 (C1649U, U2016C, A2222C, U2460C, and U3311G) were absent in the lineages TS-t25 and G-t25, whereas A1088G was only present as a polymorphism at late transfers.

There were also some mutations (between four and six per treatment) that were present at some point of the transfer series, and disappeared or decreased their representation at later transfers. The most noticeable example was the aforementioned substitution A1930G. In some cases, lost mutations were fixed in other lineages and, in other cases, substitutions that were becoming less abundant increased their frequency anew at later transfers.

Table 3. First substitutions fixed in the evolutionary lineages of bacteriophage Q β and the number of the transfer at which they reached fixation.

Lineage	First Fixed Mutations	Number of Transfer
S-t2.1	A1088G, G1371A, C2452U, U2776C	30
S-t2.2	A1088G, G1371A, A2222C, U3311G	20
S-t2.3	G1312A, U3402C	10
S-t25.1	A1088G, C1649U, U2016C, A2222C, U2460C, U3311G, G3945A	10
S-t25.2	A1088G, A2222C, U3311G	10
S-t25.3	A1088G, C1649U, A2222C, U3311G	10
TS-t25.1	G1312A	20
TS-t25.2	G1312A, A1930G	20
TS-t25.3	G1312A	30
G-t25.1	G1312A	20
G-t25.2	A1930G	20
G-t25.3	G1312A	20

2.6. Fixation of Mutations and Temperature of Evolution

Populations subjected to the progressive temperature increases evolved through a series of different environments, each of them characterized by a temperature value. The analysis of the mutations

selected in each of these environments may shed some light on the fitness effects of mutations at different temperatures. We determined which mutations were detected for the first time in the range between 38 °C and 40 °C or at temperatures above 40 °C in the lineages TS-t25 and G-t25. Most mutations appeared for the first time in the range of 38–40 °C (U1295C/G, G1371A, G1312A, C1806U, A1930G, A2187C, C2201U, U2776C, C3545U, and G3945A) (Supplementary Figure S1). Some of these substitutions were also present in most of the lineages evolving constantly at 43 °C (G1312A, U1295C/G, G1371A), suggesting that they are beneficial at all assayed temperatures. In contrast to this, C1806U, A1930G, and C2201U were absent from all the lineages evolved at a constant 43 °C, while A2187C and C3545U were present only in one. Whereas A1930G exhibited a significant propensity to be lost at later transfers, suggesting a negative effect at higher temperatures, C1806U remained in all lineages and C2201U in one (Supplementary Figure S1). Substitutions A1088G, A1778G, G2223A, and C2384T were first detected at temperatures above 40 °C (Supplementary Figure S1). All except A1778G were also present in the lineages evolving at a constant 43 °C.

3. Discussion

The results obtained in this work confirm the conclusions of previous studies showing that both the rate of change in the selective pressures and the characteristics of populations before exposure to new conditions have a great influence on the dynamics of adaptation, the achieved fitness values, the mutational pathways followed, and the repeatability of evolution [9–16]. The properties of our experimental model—an RNA phage with high mutation rates, fast replication, and large population sizes [38,49]—and of the assayed selective pressure—an environmental condition that affects most biological processes—may determine some of the particularities of our findings.

Previous studies carried out with Q β showed a discrepancy in the capacity of this virus to withstand high temperatures. Whereas the virus obtained upon expression of the infectious clone pACYCQ β [48] was unable to produce an infectious progeny at 43.6 °C [34], the virus obtained upon expression of pBRT7Q β [46,47] could replicate at 43 °C, although rendering much lower titers than at 37 °C [44]. The ability of pBRT7Q β to generate viruses able to replicate at 43 °C was probably due to the presence of two silent substitutions (C1257U and C2249U) that, together with G4A, U192C, and C2201U (also silent), allowed for adaptation to a high temperature of the virus obtained upon expression of pACYCQ β [34]. Optimal adaptation also needed the fixation of substitutions A1088G, A2748C, and U2776C, whose beneficial effects were dependent on the presence of the aforementioned set of five silent substitutions. As we have shown in this work, Q β obtained upon expression of pBRT7Q β could efficiently adapt to replication at 43 °C both when the temperature increased sharply and when it did in a more gradual manner. Thus, the presence of only C1257U and C2249U sufficed to permit virus replication at 43 °C, keeping population sizes compatible with optimal adaptation. Substitutions A1088G, U2776C, and C2201U were also detected in the evolutionary lineages analyzed in this work, all with origin in pBRT7Q β . Altogether, the results obtained with both Q β variants show the relevance of the genetic background to determine the fitness effects of mutations [3,4,50–53], in this case of those with beneficial effects at a high temperature. Since some of the mutations that facilitate adaptation of Q β to increased temperature were already present in the virus prior to high-temperature exposure, it would be expected that they could exert similar effects under other unrelated selective pressures.

The comparison of the dynamics of fitness gains in populations with the same origin but differing in the rate of temperature increase (S-t25, TS-t25, and G-t25) showed that virus propagation at temperatures above 38 °C produced fitness increases at 43 °C, suggesting that at least some of the mutations selected when the selective pressure was mild were also beneficial at a higher intensity. However, neither the fitness gains reached when the evolution experiment was stopped nor the fitness trajectories along the transfer series were similar. The lower the rate of temperature increase, the larger the number of transfers needed to duplicate the fitness of the ancestor, which agrees with the selection of mutations of smaller effect at lower intensities of the selective pressure, as predicted by theoretical models [17–20]. By transfer number 60, sudden exposure to 43 °C also led to fitness values higher than

those reached under progressive temperature changes, as observed elsewhere [31,32,35]. Differences in the fitness dynamics could be well-explained if, as mentioned above, the mutations selected in lineages subjected to different treatments had different effects on fitness and/or they accumulated at a different rate.

In good agreement with rapid increases in fitness at early transfers, the first event of fixation in lineages S-t25 involved more mutations and took place earlier than in lineages evolving under more progressive change. The fast selection of these mutations in lineages S-t25 suggests that their ancestor population probably contained some genomes that either had beneficial effects at 43 °C or could facilitate the acquisition of adaptive mutations at this temperature. The first fixed mutations were not the same in the different treatments, which could indicate that the magnitude of their effects, or even the sign of some of them, depended on the intensity of the selective pressure, something that has also been demonstrated in previous studies [12,34]. The lower fitness values reached at transfer 60 by populations evolved through the progressive treatments could also be explained if the mutations acquired during transfers carried out at lower temperatures entailed fitness costs at higher temperatures or imposed restrictions on the mutations acquired at later transfers. In line with this argument, some of the first fixed mutations in lineages S-t25, probably responsible for their large fitness increases at early transfers, were not present in the lineages evolved at a progressively increased temperature. The clearest example of a mutation with a probable fitness cost at a high temperature was A1930G, which was only present in lineages G-t25 and TS-t25 at early transfers—to disappear later as the temperature rose—and was always absent in lineages S-t25.

Populations evolving through the different treatments had different life stories, and therefore different fitness values before the first exposure to 43 °C. There was a negative linear relationship between those values and the fitness gains upon the subsequent 10 transfers at 43 °C, confirming the general observation in micro-organisms that fitter populations adapt more slowly than less fit populations [54–57]. This decrease in the rate of adaptation indicates that either better-adapted populations have already used most of the beneficial mutations available or that beneficial mutations have weaker effects on fitter genotypes. The latter is a form of epistasis known as diminishing returns epistasis [58,59], identified in previous studies, that seems to represent a generic negative coupling among beneficial mutations. Diminishing returns epistasis could explain why some of the mutations responsible for the largest fitness increases in poorly adapted populations (population Q β -t25 before propagation through the sudden treatment) were not selected when better-adapted populations were also exposed to 43 °C (lineages TS-25 at transfer number 30 and lineages G-t25 at transfer number 50).

Convergence among replicate lineages was higher for populations evolved through progressive temperature increases than for populations constantly evolved at 43 °C. Although sudden exposure to a harsh selective pressure might have caused the selection of the genomes with the most beneficial mutations that were already present in the standing genetic diversity of the ancestor population, the large number of transfers experienced at 43 °C by lineages evolving under constant conditions might have caused their divergence due to the acquisition of new mutations of small effect. Convergence between treatments was higher when lineages TS-t25 were compared with lineages G-t25 than when any of them was compared with lineages S-t25, highlighting the importance of the pattern of change in the followed evolutionary pathway.

The comparison of the dynamics of adaptation to 43 °C in lineages S-t2 and S-t25, both propagated through the sudden treatment but differing in their ancestor population, shows great differences in the rate of adaptation, which was much faster for lineages S-t25. Due to differences in their life histories, both ancestors Q β -t2 and Q β -t25 probably contain different genetic diversity, arguably lower in population Q β -t2 than in Q β -t25. The rapidity of adaptation in lineages S-t25 suggests the presence in their ancestor populations of some genomes containing mutations that facilitate adaptation to 43 °C. These genomes should be present in low frequencies, since population Q β -t25 does not have higher fitness than Q β -t2 at 43 °C. In contrast to this, adaptive mutations were probably absent from population Q β -t2 and had to be generated *de novo*, which causes the lower rate of adaptation in

lineages S-t2. This fact can also underlie the higher nucleotide diversity among replicate lineages and the larger number of transfers that is required for the fixation of the first set of mutations in lineages S-t2 as compared to lineages S-t25.

Although lineages S-t25 adapted faster than lineages S-t2, they reached similar fitness values at the end point of evolution. There was also high mutational convergence between both sets of lineages, higher than between any of them and Gt-25 or TS-t25, indicating that the pattern of change in temperature is more important than the ancestor population to determine which mutations will be selected for optimal adaptation. The main differences in the consensus sequences of lineages St-25 and S-t2 at transfer number 60 were due to mutations represented in only one evolutionary lineage. Those mutations could not have selective value and may have been fixed through hitchhiking. Most represented mutations identified in this study were also detected in previous studies in which Q β adapted to a high temperature [34,43–45]. This agreement notwithstanding, most of the mutations (12 out of 16) that were present in only one evolutionary lineage had not been previously detected during adaptation of Q β to a high temperature.

There are multiple fitness traits that can be associated with increased performance of Q β to high temperatures, among them the extracellular stability of the virus particles, the adsorption rate to the host, the time span that is required to produce an intracellular progeny, and the liberation of the new-formed viruses to the external medium. Kashiwagi et al. [60] prepared several site-directed mutants containing different sets of mutations among those selected at a high temperature, and determined some of the parameters that characterize the Q β infective cycle. All the combinations of assayed mutations increased the phage adsorption rate and decreased the latent period at 43.6 °C, which agrees with adaptation to this temperature. However, surprisingly, the same site-directed mutants displayed decreased thermal stability at 43.8 °C, which was attributed to the presence of substitution A1088G (D342G) in the A2 protein. Nucleotide D342 is placed near the surface of the capsid and has been suggested to interact with the coat protein. Thus, mutations at this position can alter the stability of the capsid [61]. Further experiments showed that stability could be restored, at least partially, through silent mutations that affect the structure of the RNA. Other substitutions that may have an effect on virus stability are C1649U (synonymous, in the coat protein), A1930G (Q195R in the A1 protein), U2016C (F224L in the A1 protein), and G3945A (G531S in the replicase). All of them were identified in an experiment designed to select an increased virus multiplication rate at 37 °C, which was associated with decreased stability at the same temperature [62]. Altogether, these results illustrate how the optimization of some fitness traits entails a cost in others, requiring the presence of additional mutations to generate an optimized genotype. Currently, the crystal structure of all the Q β proteins has been determined [61,63–66], opening the door to computational studies able to assess whether a particular mutation (or combination of mutations) influences the stability of the protein. Similar studies could be carried out with the genomic RNA.

The elucidation of the molecular mechanisms by which each of the mutations identified in this work exert its effects is a topic of interest that deserves further attention. The analysis of the mutational dynamics at the molecular level would need to track the changes in the frequency of particular substitutions through time by means of high-throughput sequencing techniques. Determination of the consensus sequences of sets of biological clones at different points of the process would be very useful to determine which associations of mutations are favored (such as U2016C, U3311G, and C3903U, which appear always together in the same lineages, suggesting that their positive effect could be dependent on their joint presence), and to determine the intra-population dynamics of adaptation. These studies are currently under progress and will help to better elucidate the relationships between the pattern of change in the environmental variables, the internal structure of populations, and the availability of adaptive pathways.

4. Materials and Methods

4.1. Virus Populations, Bacteria, and Standard Procedures for Infection

The plasmid pBRT7Q β , containing the cDNA of bacteriophage Q β cloned in the plasmid pBR322 [46,47], was used to transform *E. coli* DH5- α , which can support the expression of the virus genes and the assembly of virus particles, but cannot be infected by the virus because it lacks the F pilus. The supernatant of an overnight culture obtained from a transformed colony was used to infect *E. coli* Hfr (Hayes) in semisolid agar at a multiplicity of infection (moi) that allowed for the generation of well-separated lytic plaques. Under these conditions, there is a high probability that each lytic plaque results from the replication of a single virus for a limited number of generations, and, thus, they can be considered biological clones. The virus progeny contained in a randomly chosen lytic plaque was the origin of all the populations used in this study.

Standard infections in liquid medium were carried out in NB medium (8 g/L Nutrient Broth from Merck and 5 g/L NaCl), using fresh log-phase *E. coli* Hfr cultures with an OD₅₅₀ between 0.6 and 0.8. After 2 h of incubation at 37 °C with good aeration (250 rpm), cultures were treated with 1/20 volume of chloroform for 15 min at 37 °C with shaking (300 rpm). Virus supernatants were harvested upon centrifugation at 13,000 \times g. Virus titers were determined by a plaque assay that was carried out by mixing 300 μ L of exponential phase bacteria with 100 μ L of the phage suspension in 3.5 mL of melted top agar, which was poured onto Petri dishes containing a bottom agar layer. The phage concentration was adjusted to obtain a number of lytic plaques between 20 and 40. Titers were expressed as the number of plaque forming units (pfu) per mL of the phage suspension in the case of virus supernatants.

The biological clone isolated upon expression of pBRT7Q β was used to infect an *E. coli* Hfr culture as described above, using an moi of 0.1 pfu/cell in a volume of 1 mL (containing $\sim 10^8$ bacteria). After 2 h of incubation at 37 °C, the virus supernatant was collected, and a fraction of it was used to infect a fresh *E. coli* culture, keeping the moi around 0.1 pfu/cell. The process was repeated for a total of 25 serial transfers. Populations obtained at transfer number 2 (Q β -t2) or 25 (Q β -t25) were propagated for 60 additional transfers but, instead of at 37 °C, the replication temperature was increased according to the different patterns shown in Figure 1. When the virus yield was too low to make infections at an moi of 0.1 pfu/cell, these were performed using the virus contained in 100 μ L of the previous transfer. Evolutionary lineages were carried out in triplicate and corresponded to Q β -t25 propagated through three patterns of change: sudden (lineages S-t25.1, S-t25.2, and S-t25.3), two steps (lineages TS-t25.1, TS-t25.2, and TS-t25.3), and gradual (lineages G-t25.1, G-t25.2, and G-t25.3); and to Q β -t2 propagated only through the sudden pattern (lineages S-t2.1, S-t2.2, and S-t2.3).

4.2. Fitness Determinations

Duplicate liquid cultures containing 10^8 bacteria growing in exponential phase were inoculated with 10^4 pfu in a final volume of 1 mL. After two hours of incubation at 43 °C, the virus supernatants were collected as described above and titrated to estimate the virus yield, which was used as a surrogate of fitness. Previous assays showed that bacteriophage Q β grew exponentially under these conditions. Absolute fitness was expressed as the number of doublings per two hours, and was calculated as $\log_2[(N_f - N_0)/N_0]$, where N_0 is the initial input of virus and N_f is the number of progeny pfu. Relative fitness values were obtained by dividing the absolute values by the one of the corresponding ancestors (Q β -t2 or Q β -t25) determined in the same assay.

4.3. Determination of the Consensus Sequences of Virus Populations

RNA was extracted by treatment of the virus supernatants with 1/10 vol. 10% SDS, 1/10 vol. 50% β -mercaptoethanol, and 1 μ L of RNAsin for 20 min at room temperature with low shaking. After 5 min in ice, the sample was extracted with cold-water-saturated phenol and the upper phase was precipitated with ethanol following standard procedures. RNA samples were resuspended in RNase-free water and stored at -80 °C. RNAs were used for cDNA synthesis with the avian myeloblastosis virus

reverse transcriptase (Promega), followed by PCR amplification using Expand high-fidelity DNA polymerase (Roche). The oligonucleotide primers used for PCR have been described previously [67,68]. PCR products were column-purified (Qiagen) and subjected to standard Sanger sequencing using Big Dye Chemistry with an automated sequencer (Applied Biosystems; Perkin Elmer). Sequences were assembled with SeqMan Pro (DNASTAR Lasergene 12 Core Suite) and aligned to the sequence of the cDNA of bacteriophage Q β cloned in the plasmid pBR322 using the program Clustal X2.1 (Multiple Alignment Mode and default parameters). Mutations relative to this sequence were identified using BioEdit Sequence Alignment Editor. To make a more thorough analysis, the mutations, either fixed or polymorphic, that were detected in each evolutionary lineage were visually inspected in the chromatograms of the other lineages to evaluate their possible presence as polymorphisms not recognized by the sequence analysis programs.

Supplementary Materials: The following are available online at <http://www.mdpi.com/2076-0817/8/2/80/s1>, **Figure S1:** Mutations present in the consensus sequences determined every 10 transfers for the Q β evolutionary lineages propagated through different patterns of temperature increase. The three squares below each mutation in each of the six rows shown in the figure correspond to the three replicate lineages at transfer number 10 (T10), 20 (T20), 30 (T30), 40 (T40), 50 (T50), and 60 (T60). A white square means the absence of the mutation. In colored squares, the intensity of the color indicates the relative amount of the mutated nucleotide according to the code shown in the figure. **Figure S2:** Dynamics of the accumulation of mutations in the consensus sequences of the Q β evolutionary lineages. In each treatment, the blue line represents the evolutionary lineage 1, the orange line the lineage 2, and the grey line the lineage 3. The number of mutations includes both fixed and polymorphic mutations.

Author Contributions: Conceptualization, E.L. and S.M.; methodology, P.S., E.L. and S.M.; software, P.S. and E.L.; formal analysis, P.S. and E.L.; investigation, P.S.; writing—original draft preparation, E.L.; writing—review and editing, E.L. and S.M.; supervision, E.L.; project administration, S.M.; funding acquisition, S.M. and E.L. All authors have read and approved the final manuscript.

Funding: This research was funded by the Spanish Ministerio de Ciencia, Innovación y Universidades, grant number FIS2017-89773-P.

Acknowledgments: We thank M. Arribas and E. Acosta for technical assistance.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Elena, S.F.; Lenski, R.E. Evolution experiments with microorganisms: The dynamics and genetic bases of adaptation. *Nat. Rev. Genet.* **2003**, *4*, 457–469. [[CrossRef](#)] [[PubMed](#)]
2. Kawecki, T.J.; Lenski, R.E.; Ebert, D.; Hollis, B.; Olivieri, I.; Whitlock, M.C. Experimental evolution. *Trends Ecol. Evol.* **2012**, *27*, 547–560. [[CrossRef](#)] [[PubMed](#)]
3. Lang, G.I.; Desai, M.M. The spectrum of adaptive mutations in experimental evolution. *Genomics* **2014**, *104*, 412–416. [[CrossRef](#)] [[PubMed](#)]
4. Jerison, E.R.; Desai, M.M. Genomic investigations of evolutionary dynamics and epistasis in microbial evolution experiments. *Curr. Opin. Genet. Dev.* **2015**, *35*, 33–39. [[CrossRef](#)] [[PubMed](#)]
5. Bailey, S.F.; Bataillon, T. Can the experimental evolution programme help us elucidate the genetic basis of adaptation in nature? *Mol. Ecol.* **2016**, *25*, 203–218. [[CrossRef](#)] [[PubMed](#)]
6. Elena, S.F.; Sanjuán, R. Virus Evolution: Insights from an Experimental Approach. *Annu. Rev. Ecol. Evol. Syst.* **2007**, *38*, 27–52. [[CrossRef](#)]
7. Collins, S. Many Possible Worlds: Expanding the Ecological Scenarios in Experimental Evolution. *Evol. Biol.* **2011**, *38*, 3–14. [[CrossRef](#)]
8. Barrick, J.E.; Lenski, R.E. Genome dynamics during experimental evolution. *Nat. Rev. Genet.* **2013**, *14*, 827–839. [[CrossRef](#)]
9. Collins, S.; de Meaux, J. Adaptation to different rates of environmental change in *chlamydomonas*. *Evolution* **2009**, *63*, 2952–2965. [[CrossRef](#)]
10. Toprak, E.; Veres, A.; Michel, J.B.; Chait, R.; Hartl, D.L.; Kishony, R. Evolutionary paths to antibiotic resistance under dynamically sustained drug selection. *Nat. Genet.* **2012**, *44*, 101–105. [[CrossRef](#)]

11. Gonzalez, A.; Bell, G. Evolutionary rescue and adaptation to abrupt environmental change depends upon the history of stress. *Philos. Trans. R. Soc. Lond. B* **2012**, *368*, 20120079. [[CrossRef](#)] [[PubMed](#)]
12. Lindsey, H.A.; Gallie, J.; Taylor, S.; Kerr, B. Evolutionary rescue from extinction is contingent on a lower rate of environmental change. *Nature* **2013**, *494*, 463–467. [[CrossRef](#)] [[PubMed](#)]
13. Morley, V.J.; Mendiola, S.Y.; Turner, P.E. Rate of novel host invasion affects adaptability of evolving RNA virus lineages. *Proc. R. Soc. B* **2015**, *282*, 20150801. [[CrossRef](#)] [[PubMed](#)]
14. Gorter, F.A.; Aarts, M.M.; Zwaan, B.J.; de Visser, J.A.G.M. Dynamics of Adaptation in Experimental Yeast Populations Exposed to Gradual and Abrupt Change in Heavy Metal Concentration. *Am. Nat.* **2016**, *187*, 110–119. [[CrossRef](#)] [[PubMed](#)]
15. Morley, V.J.; Turner, P.E. Dynamics of molecular evolution in RNA virus populations depend on sudden versus gradual environmental change. *Evolution* **2017**, *71*, 872–883. [[CrossRef](#)] [[PubMed](#)]
16. Gorter, F.A.; Derks, M.F.L.; van den Heuvel, J.; Aarts, M.G.M.; Zwaan, B.J.; de Ridder, D.; de Visser, J.A.G.M. Genomics of Adaptation Depends on the Rate of Environmental Change in Experimental Yeast Populations. *Mol. Biol. Evol.* **2017**, *34*, 2613–2626. [[CrossRef](#)] [[PubMed](#)]
17. Bello, Y.; Waxman, D. Near-periodic substitution and the genetic variance induced by environmental change. *J. Theor. Biol.* **2006**, *239*, 152–160. [[CrossRef](#)] [[PubMed](#)]
18. Kopp, M.; Hermisson, J. Adaptation of a Quantitative Trait to a Moving Optimum. *Genetics* **2007**, *176*, 715–719. [[CrossRef](#)] [[PubMed](#)]
19. Kopp, M.; Hermisson, J. The Genetic Basis of Phenotypic Adaptation II: The Distribution of Adaptive Substitutions in the Moving Optimum Model. *Genetics* **2009**, *183*, 1453–1476. [[CrossRef](#)]
20. Kopp, M.; Hermisson, J. The Genetic Basis of Phenotypic Adaptation I: Fixation of Beneficial Mutations in the Moving Optimum Model. *Genetics* **2009**, *182*, 233–249. [[CrossRef](#)]
21. Rokyta, D.R.; Joyce, P.; Caudle, S.B.; Wichman, H.A. An empirical test of the mutational landscape model of adaptation using a single-stranded DNA virus. *Nat. Genet.* **2005**, *37*, 441–444. [[CrossRef](#)] [[PubMed](#)]
22. Eyre-Walker, A.; Keightley, P.D. The distribution of fitness effects of new mutations. *Nat. Rev. Genet.* **2007**, *8*, 610–618. [[CrossRef](#)] [[PubMed](#)]
23. Bons, E.; Bertels, F.; Regoes, R.R. Estimating the mutational fitness effects distribution during early HIV infection. *Virus Evol.* **2018**, *4*, vey029. [[CrossRef](#)] [[PubMed](#)]
24. Bürger, R.; Lynch, M. Evolution and extinction in a changing environment: A quantitative-genetic analysis. *Evolution* **1995**, *49*, 151–163. [[CrossRef](#)] [[PubMed](#)]
25. Bell, G.; Collins, S. Adaptation, extinction and global change: Adaptation, extinction and global change. *Evol. Appl.* **2008**, *1*, 3–16. [[CrossRef](#)] [[PubMed](#)]
26. Gomulkiewicz, R.; Houle, D. Demographic and Genetic Constraints on Evolution. *Am. Nat.* **2009**, *174*, E218–E229. [[CrossRef](#)] [[PubMed](#)]
27. Kopp, M.; Matuszewski, S. Rapid evolution of quantitative traits: Theoretical perspectives. *Evol. Appl.* **2014**, *7*, 169–191. [[CrossRef](#)] [[PubMed](#)]
28. Yubero, P.; Manrubia, S.; Aguirre, J. The space of genotypes is a network of networks: Implications for evolutionary and extinction dynamics. *Sci. Rep.* **2017**, *7*, 13813. [[CrossRef](#)] [[PubMed](#)]
29. Schenk, M.F.; Szendro, I.G.; Salverda, M.L.M.; Krug, J.; de Visser, J.A.G.M. Patterns of Epistasis between Beneficial Mutations in an Antibiotic Resistance Gene. *Mol. Biol. Evol.* **2013**, *30*, 1779–1787. [[CrossRef](#)] [[PubMed](#)]
30. Schoustra, S.; Hwang, S.; Krug, J.; de Visser, J.A.G.M. Diminishing-returns epistasis among random beneficial mutations in a multicellular fungus. *Proc. R. Soc. B* **2016**, *283*, 20161376. [[CrossRef](#)] [[PubMed](#)]
31. Perron, G.G.; Gonzalez, A.; Buckling, A. The rate of environmental change drives adaptation to an antibiotic sink. *J. Evol. Biol.* **2008**, *21*, 1724–1731. [[CrossRef](#)] [[PubMed](#)]
32. Bell, G.; Gonzalez, A. Adaptation and Evolutionary Rescue in Metapopulations Experiencing Environmental Deterioration. *Science* **2011**, *332*, 1327–1330. [[CrossRef](#)] [[PubMed](#)]
33. Hao, Y.-Q.; Brockhurst, M.A.; Petchey, O.L.; Zhang, Q.-G. Evolutionary rescue can be impeded by temporary environmental amelioration. *Ecol. Lett.* **2015**, *18*, 892–898. [[CrossRef](#)] [[PubMed](#)]
34. Kashiwagi, A.; Sugawara, R.; Sano Tsushima, F.; Kumagai, T.; Yomo, T. Contribution of Silent Mutations to Thermal Adaptation of RNA Bacteriophage Q β . *J. Virol.* **2014**, *88*, 11459–11468. [[CrossRef](#)] [[PubMed](#)]
35. Baym, M.; Lieberman, T.D.; Kelsic, E.D.; Chait, R.; Gross, R.; Yelin, I.; Kishony, R. Spatiotemporal microbial evolution on antibiotic landscapes. *Science* **2016**, *353*, 1147–1151. [[CrossRef](#)] [[PubMed](#)]

36. Singhal, S.; Leon Guerrero, C.M.; Whang, S.G.; McClure, E.M.; Busch, H.G.; Kerr, B. Adaptations of an RNA virus to increasing thermal stress. *PLoS ONE* **2017**, *12*, e0189602. [[CrossRef](#)] [[PubMed](#)]
37. Olsthoorn, R.; van Duin, J. Bacteriophages with ssRNA. In *eLS*; John Wiley & Sons, Ltd.: Chichester, UK, 2011. [[CrossRef](#)]
38. Bradwell, K.; Combe, M.; Domingo-Calap, P.; Sanjuán, R. Correlation Between Mutation Rate and Genome Size in Riboviruses: Mutation Rate of Bacteriophage Q β . *Genetics* **2013**, *195*, 243–251. [[CrossRef](#)] [[PubMed](#)]
39. Moya, A.; Holmes, E.C.; González-Candelas, F. The population genetics and evolutionary epidemiology of RNA viruses. *Nat. Rev. Microbiol.* **2004**, *2*, 279–288. [[CrossRef](#)] [[PubMed](#)]
40. Lázaro, E. RNA Viruses: Control, Mutagenesis and Extinction. In *eLS*; John Wiley & Sons, Ltd.: Chichester, UK, 2011; p. a0023276, ISBN 978-0-470-01617-6.
41. Domingo, E.; Sheldon, J.; Perales, C. Viral Quasispecies Evolution. *Microbiol. Mol. Biol. Rev.* **2012**, *76*, 159–216. [[CrossRef](#)] [[PubMed](#)]
42. Guzella, T.S.; Dey, S.; Chelo, I.M.; Pino-Querido, A.; Pereira, V.F.; Proulx, S.R.; Teotónio, H. Slower environmental change hinders adaptation from standing genetic variation. *PLoS Genet.* **2018**, *14*, e1007731. [[CrossRef](#)]
43. Arribas, M.; Kubota, K.; Cabanillas, L.; Lázaro, E. Adaptation to Fluctuating Temperatures in an RNA Virus Is Driven by the Most Stringent Selective Pressure. *PLoS ONE* **2014**, *9*, e100940. [[CrossRef](#)] [[PubMed](#)]
44. Arribas, M.; Cabanillas, L.; Kubota, K.; Lázaro, E. Impact of increased mutagenesis on adaptation to high temperature in bacteriophage Q β . *Virology* **2016**, *497*, 163–170. [[CrossRef](#)] [[PubMed](#)]
45. Lázaro, E.; Arribas, M.; Cabanillas, L.; Román, I.; Acosta, E. Evolutionary adaptation of an RNA bacteriophage to the simultaneous increase in the within-host and extracellular temperatures. *Sci. Rep.* **2018**, *8*, 8080. [[CrossRef](#)] [[PubMed](#)]
46. Taniguchi, T.; Palmieri, M.; Weissmann, C. Q β DNA-containing hybrid plasmids giving rise to Q β phage formation in the bacterial host. *Nature* **1978**, *274*, 223–228. [[CrossRef](#)] [[PubMed](#)]
47. Barrera, I.; Schuppli, D.; Sogo, J.M.; Weber, H. Different Mechanisms of Recognition of Bacteriophage Q β Plus and Minus Strand RNAs by Q β Replicase. *J. Mol. Biol.* **1993**, *232*, 512–521. [[CrossRef](#)] [[PubMed](#)]
48. Inomata, T.; Kimura, H.; Hayasaka, H.; Shiozaki, A.; Fujita, Y.; Kashiwagi, A. Quantitative comparison of the RNA bacteriophage Q β infection cycle in rich and minimal media. *Arch. Virol.* **2012**, *157*, 2163–2169. [[CrossRef](#)] [[PubMed](#)]
49. Callanan, J.; Stockdale, S.; Shkoporov, A.; Draper, L.; Ross, R.; Hill, C. RNA Phage Biology in a Metagenomic Era. *Viruses* **2018**, *10*, 386. [[CrossRef](#)] [[PubMed](#)]
50. Weinreich, D.M.; Watson, R.A.; Chao, L. Perspective: Sign epistasis and genetic constraint on evolutionary trajectories. *Evolution* **2005**, *59*, 1165–1174. [[CrossRef](#)]
51. Parera, M.; Martinez, M.A. Strong Epistatic Interactions within a Single Protein. *Mol. Biol. Evol.* **2014**, *31*, 1546–1553. [[CrossRef](#)] [[PubMed](#)]
52. Lyons, D.; Luring, A. Mutation and Epistasis in Influenza Virus Evolution. *Viruses* **2018**, *10*, 407. [[CrossRef](#)] [[PubMed](#)]
53. Bernet, G.P.; Elena, S.F. Distribution of mutational fitness effects and of epistasis in the 5' untranslated region of a plant RNA virus. *BMC Evol. Biol.* **2015**, *15*, 274. [[CrossRef](#)] [[PubMed](#)]
54. Wiser, M.J.; Ribick, N.; Lenski, R.E. Long-Term Dynamics of Adaptation in Asexual Populations. *Science* **2013**, *342*, 1364–1367. [[CrossRef](#)] [[PubMed](#)]
55. Perfeito, L.; Sousa, A.; Bataillon, T.; Gordo, I. Rates of fitness decline and rebound suggest pervasive epistasis: Pervasive epistasis in fitness decline and rebound. *Evolution* **2014**, *68*, 150–162. [[CrossRef](#)] [[PubMed](#)]
56. Kryazhimskiy, S.; Rice, D.P.; Jerison, E.R.; Desai, M.M. Global epistasis makes adaptation predictable despite sequence-level stochasticity. *Science* **2014**, *344*, 1519–1522. [[CrossRef](#)] [[PubMed](#)]
57. Couce, A.; Tenaillon, O.A. The rule of declining adaptability in microbial evolution experiments. *Front. Genet.* **2015**, *6*, 99. [[CrossRef](#)] [[PubMed](#)]
58. Chou, H.-H.; Chiu, H.-C.; Delaney, N.F.; Segre, D.; Marx, C.J. Diminishing Returns Epistasis Among Beneficial Mutations Decelerates Adaptation. *Science* **2011**, *332*, 1190–1192. [[CrossRef](#)] [[PubMed](#)]
59. Khan, A.I.; Dinh, D.M.; Schneider, D.; Lenski, R.E.; Cooper, T.F. Negative Epistasis Between Beneficial Mutations in an Evolving Bacterial Population. *Science* **2011**, *332*, 1193–1196. [[CrossRef](#)] [[PubMed](#)]

60. Kashiwagi, A.; Kadoya, T.; Kumasaka, N.; Kumagai, T.; Tsushima, F.S.; Yomo, T. Influence of adaptive mutations, from thermal adaptation experiments, on the infection cycle of RNA bacteriophage Q β . *Arch. Virol.* **2018**, *163*, 2655–2662. [[CrossRef](#)] [[PubMed](#)]
61. Rumnieks, J.; Tars, K. Crystal Structure of the Maturation Protein from Bacteriophage Q β . *J. Mol. Biol.* **2017**, *429*, 688–696. [[CrossRef](#)] [[PubMed](#)]
62. García-Villada, L.; Drake, J.W. Experimental selection reveals a trade-off between fecundity and lifespan in the coliphage Q β . *Open Biol.* **2013**, *3*, 130043. [[CrossRef](#)]
63. Rumnieks, J.; Tars, K. Crystal structure of the read-through domain from bacteriophage Q β A1 protein. *Protein Sci.* **2011**, *20*, 1707–1712. [[CrossRef](#)] [[PubMed](#)]
64. Rumnieks, J.; Tars, K. Crystal structure of the bacteriophage Q β coat protein in complex with the RNA operator of the replicase gene. *J. Mol. Biol.* **2014**, *426*, 1039–1049. [[CrossRef](#)] [[PubMed](#)]
65. Takeshita, D.; Tomita, K. Assembly of Q β viral RNA polymerase with host translational elongation factors EF-Tu and -Ts. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 15733–15738. [[CrossRef](#)] [[PubMed](#)]
66. Kidmose, R.T.; Vasiliev, N.N.; Chetverin, A.B.; Andersen, G.R.; Knudsen, C.R. Structure of the Q β replicase, an RNA-dependent RNA polymerase consisting of viral and host proteins. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 10884–10889. [[CrossRef](#)] [[PubMed](#)]
67. Arribas, M.; Cabanillas, L.; Lázaro, E. Identification of mutations conferring 5-azacytidine resistance in bacteriophage Q β . *Virology* **2011**, *417*, 343–352. [[CrossRef](#)] [[PubMed](#)]
68. Cabanillas, L.; Arribas, M.; Lázaro, E. Evolution at increased error rate leads to the coexistence of multiple adaptive pathways in an RNA virus. *BMC Evol. Biol.* **2013**, *13*, 11. [[CrossRef](#)] [[PubMed](#)]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).